Universidade de Aveiro
2016

Departamento de Eletrónica,
Telecomunicações e Informática

Helder Moitalta
Capelo

Novas Soluções de Mobilidade em Redes 3GPP

New Mobility Solutions in 3GPP Networks

Universidade de Aveiro 2016

Departamento de Eletrónica, Telecomunicações e Informática

Helder Moitalta Capelo

Novas Soluções de Mobilidade em Redes 3GPP

New Mobility Solutions in 3GPP Networks

"*To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science.*"

— Albert Einstein

Universidade de Aveiro
2016

Departamento de Eletrónica,
Telecomunicações e Informática

Helder Moitalta Capelo

Novas Soluções de Mobilidade em Redes 3GPP

New Mobility Solutions in 3GPP Networks

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Eletrónica e Telecomunicações, realizada sob a orientação científica do Professor Doutor Rui L. Aguiar, Professor catedrático do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro, e do Doutor Daniel Corujo, Professor adjunto convidado da Escola Superior de Tecnologia e Gestão de Águeda da Universidade de Aveiro.

Bolsa de investigação no âmbito
do projeto IM3W.

Dedico este trabalho aos meus pais e a todos os que partilharam comigo este caminho.

o júri / the jury

presidente / president	Professor Doutor José Alberto Gouveia Fonseca Professor Associado, Universidade de Aveiro
vogais / examiners committee	Doutor Pedro Miguel Naia Neves Consultor Sénior, Altice Labs
	Professor Doutor Rui Luís Andrade Aguiar Professor Catedrático, Universidade de Aveiro

agradecimentos / acknowledgements

Agradeço aos meus orientadores, Professor Doutor Rui L. Aguiar e Doutor Daniel Corujo, pelo inefável apoio e pelo papel fundamental que desempenharam no desenvolvimento desta dissertação.

Á equipa do projeto IM3W da Altice Labs (PT Inovação e Sistemas) e do ATNoG - Aveiro Telecommunications and Networking Group deixo um agradecimento pelo que contribuiram para este trabalho.

Agradeço ainda aos colaboradores do Instituto de Telecomunicações de Aveiro e em especial ao Rainho por todo o apoio prestado.

Palavras Chave

3GPP, LTE, ANDSF, UPCON, Congestão, Mobilidade, Wi-Fi.

Resumo

As redes moveis tornaram-se ubíquas e a introdução de dispositivos moveis com capacidade de ligação à Internet e necessidade de grande largura de banda têm vindo a exigir da rede uma crescente capacidade de resposta, o que implica grandes investimentos na infraestrutura. Para maximizar o retorno deste investimento os operadores recorrem a técnicas de gestão da congestão da rede. Para isso podem recorrer ao *Access Network and Discovery Function* (ANDSF) que permite aproveitar o facto de estes dispositivos possuírem capacidade que se ligarem tanto a redes moveis *3rd Generation Partnership Project* (3GPP) como a redes Wi-Fi, o que possibilita um balanceamento do tráfego entre estes dois tipos de acesso.
Esta dissertação aborda a problemática da gestão da congestão na rede móvel do operador recorrendo ao mecanismo *User Plane Congestion Management* (UPCON) recentemente standardizado pelo 3GPP e propõe que ele seja usado para melhorar o funcionamento do ANDSF. Os resultados obtidos em ambiente de simulação revelam que esta solução favorece a qualidade de serviço ao utilizador com um impacto reduzido na infraestrutura do operador.

Keywords 3GPP, LTE, ANDSF, UPCON, Congestion, Mobility, Wi-Fi.

Abstract

Mobile Networks have become ubiquous and the introduction of mobile devices with internet conection and large bandwidth requirements have been demanding a growing network capacity which implies large infrestructure investiments. To maximize the return of this investiment operators use network congestion management technologies. They can resort to the Access Network and Discovery Function (ANDSF) which takes advantage of the fact that these devices have the ability to connect either to 3rd Generation Partnership Project (3GPP) networks or to Wi-Fi networks which enables traffic load ballance between these types of access.

This dissertation approches the issue of congestion management in an operator mobile network using the 3GPP's recently standardized User Plane Congestion Management (UPCON) mechanism and proposes that it should be used to improve the ANDSF. The results from the simulation reveal that this solution improves the user experience with minimal impact on the operator infraestructure.

Contents

List of Figures

List of Tables

List of Acronyms

3G	3rd Generation Mobile Network
3GPP	3rd Generation Partnership Project
ANDSF	Access Network and Discovery Function
ANDSF-C	ANDSF Client
ANDSF-S	ANDSF Server
ANQP	Access Network Query Protocol
AP	Access Point
APN	Access Point Name
ARA	Aggregated RUCI Report Answer
ARPU	Average Revenue Per User
ARR	Aggregated RUCI Report Request
AS	Access Stratum
CAPEX	Capital Expenditures
CN	Core Network
CSV	Comma Separated Values
CT	Core Network & Terminals
DRA	Diameter Routing Agent
ECGI	E-UTRAN Cell Global Identifier
eNB	eNodeB
EPC	Evolved Packet Core
EPS	Evolved Packet System
E-SMLC	Evolved SMLC
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved UTRAN
FDD	Frequency Division Duplex
FQDN	Fully Qualified Domain Name
GAS	Generic Advertisement Service
GERAN	GSM EDGE Radio Access Networks

GPRS	General Packet Radio Service
GUI	Graphical User Interface
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
IARP	Inter-APN Routing Policy
IEEE	Institute of Electrical and Electronics Engineers
IFOM	IP Flow Mobility
IM3W	Integração e Mobilidade em redes 3GPP-WiFi
IMSI	International Mobile Subscriber Identity
IMT	International Mobile Telecommunication
ISMP	Inter System Mobility Policy
ISRP	Inter System Routing Policy
IT	Instituto de Telecomunicações
LIPA	Local IP Access
LTE	Long Term Evolution
LWA	LTE-WLAN Aggregation
LWIP	LTE WLAN Radio Level Integration with IPsec Tunnel
MAC	Media Access Control
MAPCON	Multi Access PDN Connectivity
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
MNO	Mobile Network Operator
MO	Management Object
MUA	Modify Uecontext Answer
MUR	Modify Uecontext Request
NAS	Non-Access Stratum
NBIFOM	Network-Based IP Flow Mobility
NRA	Non-Aggregated RUCI Report Answer
NRR	Non-Aggregated RUCI Report Request
OAM	Operations, Administration and Maintenance
OFDM	Orthogonal Frequency Division Multiple Access
PCC	Policy and Charging Control
PCEF	Policy and Charging Enforcing Function
PCRF	Policy and Charging Rules Function
PDN	Packet Data Network
PGW	PDN Gateway
PHY	Physical Layer
PLMN	Public Land Mobile Network

PT	Portugal Telecom
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RCAF	RAN Congestion Awareness Function
ROI	Return of Investment
RUCI	RAN User Plane Congestion Information
SA	Service & Systems Aspects
SAI	Service Area Identifier
SC-FDMA	Single Carrier Frequency Division Multiple Access
SCTP	Stream Control Transmission Protocol
SGSN	Serving GPRS Support Node
SGW	Serving Gateway
SIP	Session Initiation Protocol
SIPTO	Selected IP Traffic Offload
TAI	Tracking Area Identity
TDD	Time Division Duplex
TDF	Traffic Detection Function
TFT	Traffic Flow Template
TLV	Type, Length and Value (encoding)
TSG	Technical Specification Group (3GPP)
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
UPCON	User Plane Congestion Management
UTRAN	Universal Terrestrial Radio Access Network
VoIP	Voice over IP
VoLTE	Voice over LTE
WG	Work Groups (3GPP)
WLAN	Wireless Local Area Network
WLANSP	WLAN Selection Policy
XML	Extensible Markup Language

CHAPTER 1

INTRODUCTION

Cellular networks have become ubiquitous around the globe as the number of mobile subscriptions reaches 7.3 billion nowadays [1]. Mobile Network Operators are expanding their infrastructure even to rural areas to cover increasing demand for voice and mobile data services, as users are so eager for service that they are even willing to build their networks with open-source hardware [2]. The use of devices like smartphones, tablets or routers with a cellular data connection to access the Internet has been growing rapidly motivated by Web 2.0 and the rise of social media and video streaming [1]. The number of users and devices accessing the cellular networks and the volume of data demand are going to keep rising over the next few years [3] [1] [4] as new technologies and a plethora of smart devices will continue to contribute to the economic and social development across the world, changing the way we interact with each-other as envisioned by some of the industry leaders[5].

"Always best connected" is the concept driving current and future mobile network development [6] [7]. Clients want to be able to connect to their operator's network and use services according to their needs. Its up to the operator to fulfill these demands. Extending a cellular network is not only very expensive but also complex as parts from various vendors are used and new and legacy hardware have to work side-by-side. Network planning and optimization are essential as the operator wants to maximize its Return of Investment (ROI) and get good Average Revenue Per User (ARPU) levels but must guarantee a good QoS and Quality of Experience (QoE) to the user in order to get a reduced subscriber churn rate[1].

Collecting network congestion information allows for the application of policies to minimize the impact on users but it is important that this procedure is standardized so that all vendors' products are compatible. With this goal the 3rd Generation Partnership Project[2] envisioned the User Plane Congestion Management (UPCON) feature and has recently completed most of the standardization work for its Release 13 [8] and vendors may now update their products to comply with the specification.

3GPP has been promoting the collaboration of its cellular networks with IEEE 802.11 based networks such as Wi-Fi or WiMax since Release 8 with the ANDSF feature, and more recently with Network-Based IP Flow Mobility (NBIFOM) and LTE-WLAN Aggregation (LWA) or LTE WLAN Radio Level Integration with IPsec Tunnel (LWIP). The deployment of a carrier Wi-Fi network is essential for congestion management as most mobile user devices support dual wireless connectivity

[1] http://www.investopedia.com/terms/c/churnrate.asp
[2] http://www.3gpp.org/

(3G/4G and Wi-Fi, despite the fact that their simultaneous utilization is still not widely supported in the more popular mobile operating systems), therefore it enables the offloading of Internet traffic to a Wi-Fi AP coping with the large user bandwidth demands and reducing 3G/4G cellular load.

This work intends to explore and contribute with impact analysis in regards to the enhancement of mobility management assistance technologies in congested 3GPP mobile networks. An architecture is proposed where the recently standardized UPCON network function, which under its current standard is used for congestion management at the LTE Core Network level, is extended to inter-work with the Access Network and Discovery Function (ANDSF). This solution improves the ANDSF mechanism as it enables the use of traffic offloading to Wi-Fi APs responding to the current level of network congestion. Using a simulation environment this concept was tested. The results from the simulation reveal improvement in user experience with minimal impact on the operator infrastructure.

1.1 MOTIVATION AND OBJECTIVES

When a MNO offloads traffic from their cellular network to a Wi-Fi AP interworking challenges arise, as this technology has different QoS characteristics and offloading certain types of traffic for some types of users may not be the best option for operators as it may have impact on user experience. Focusing on the ANDSF which is used by operators to provide their users with network selection policies, in order to grantee maximum user satisfaction, the policy rules engine shall take information regarding network congestion into account. Considering that under the current standard, the ANDSF rules engine receives no such information and that 3GPP has recently standardized a congestion management mechanism (UPCON) the main objectives are:

1. Study the state of art on mobile network congestion management in 3GPP networks focusing on the ANDSF and UPCON standards;
2. Perform a comprehensive analysis of an existing simulation on the deployment of ANDSF in a city environment;
3. Implement a simulation extending existing previous work developing a scenario where it is possible to determine the advantage of using the UPCON information reporting to enhance the ANDSF rules engine through signaling, access type and congestion metrics;

1.2 METHODOLOGY

With the goal set on implementing a simulation able to evaluate the benefits of extending the ANDSF mechanism with cell congestion information provided via the UPCON feature, a comprehensive study of both 3GPP specifications was performed. Furthermore, the study also focused on the 3GPP LTE network architecture and other traffic management mechanisms able to facilitate offloading Internet traffic to Wi-Fi networks.

Then, focusing on the core of the work, a deep analysis of the previous simulation work done under the SMCon[3] project was performed to familiarize with the environment and identify the possible enhancements to the implemented ANDSF rules engine as well as new evaluation metrics. The results of the simulation were also evaluated with metrics such as signaling events, amount of data exchanged and access type per user and use case. With this knowledge, a new simulation was developed which extended the previous one implementing a congestion reporting method, the congestion event trigger and metrics for congestion in the 3GPP cell and congested Wi-Fi AP usage. Finally, using the developed work, several simulations were run testing different scenarios and policies and the output metrics were evaluated.

The simulation results reveal that the proposed architecture leads to significant improvements in general user experience, and according to the selected policy, permits zero bad service to Premium users, all with minimal impact on the operator infrastructure.

1.3 CONTRIBUTIONS

This work was performed in the context of the Integração e Mobilidade em redes 3GPP-WiFi (IM3W)[4] project, a collaboration of Instituto de Telecomunicações (IT) polo de Aveiro and Altice Labs (formerly Portugal Telecom (PT) Inovação e Sistemas). This work consisted on studying several flow mobility mechanisms and ultimately focusing on the UPCON specification. Scenarios were proposed and simulated to evaluate the importance of UPCON in enhancing the ANDSF rules engine.

1.4 DISSERTATION LAYOUT

The remainder of this dissertation is organized as follows: Chapter 2 introduces the drivers behind the need for congestion management mechanisms in 3GPP networks, including a brief description of the main standardization entities in the field and presenting a brief description of the 3GPP LTE architecture focusing on the essential access and core network elements. An overview of several standardized 3GPP network management features is also presented in this chapter and is concluded in Chapter 3 with a detailed overview on UPCON. Chapter 4 details an existing simulator developed under the SMCon project which implemented an ANDSF Client-Server communication based on static policies and served as the basis for this work. The results of said project are also analyzed. The main task of this work is presented in Chapter 5, where the results of the simulator introduced in the previous chapter are extended to support congestion management reporting with UPCON enabling ANDSF dynamic policies to manage congestion at the cell presenting the new simulation results. The dissertation concludes with Chapter 6 with the evaluation of this work and indicating as well the future work.

[3] http://atnog.av.it.pt/content/smcon-smart-mobile-user-connectivity
[4] http://atnog.av.it.pt/content/im3w-integra-o-e-mobilidade-em-redes-3gpp-wifi

CHAPTER 2

Congestion Management in Mobile Networks

2.1 Drivers for congestion management

The rising numbers of users and mobile data usage versus the costs of extending the infrastructure forces MNOs to adopt congestion management technologies on their networks. Moreover, for operators, it is essential to provide their clients with fast and smooth access to the Internet to keep them satisfied.

2.1.1 Mobile data usage

The phenomenal success of the Apple iPhone, introduced in 2008, completely revolutionized the mobile phone industry. Other manufacturers were forced to follow Apple's success and user demand for smartphones and introduce more devices, such as those using Google's Android[1]. The number of shipped units has been rising ever since by double-digits motivated not only by their affordability but also by the need of users to be constantly connected to the Internet. Gartner predicts a slow down in the growth of smartphone sales starting in 2016 [9], but even so, close to 2 billion devices will be sold each year up to 2019. According to Ericsson [1] in the third quarter of 2015 (Q3 2015) 75% of the sold mobile phones were smartphones and 45% of the more than 7 billion global mobile subscriptions were associated with smartphones. Also, mobile broadband subscriptions have passed 4 billion by Q3 2015 and will reach 7.7 billion by 2021 (85% of the total) and will account for the majority of broadband accesses. Smartphones have become the preferred device for access to mobile broadband largely surpassing tablets and laptops. Polls show that that some users are dropping their fixed access and favor mobile broadband [10].

The amount of mobile data traffic was 65 percent higher in Q3 2015 compared to the previous year and reached an average of 5.3 EB/month but the prediction is that by 2021 the average will be 51 EB/month [1]. This enormous use of data is mainly caused by video consumption in these devices and it is due to [11]:

[1] https://www.android.com/

- Larger screens and higher resolution which enable better picture quality for streamed video;
- News and social media applications as well as advertisements increasingly rely on video content;
- Video streaming from YouTube[2] and Netflix[3] is growing (in many mobile networks today 50–70% of video traffic is from YouTube);
- Video on-demand services (VoD) such as HBO[4] and Netflix cause longer viewing times;
- Streaming on-demand and time-shifted content as reaching all consumer segments;
- TV and video content are being consumed everywhere increasingly over mobile networks and on multiple devices including smartphones;

Smartphone traffic as seen in Figure 2.1 is expected to grow 11 times from 2015 to 2021 and represent around 90 % of mobile data traffic [1].

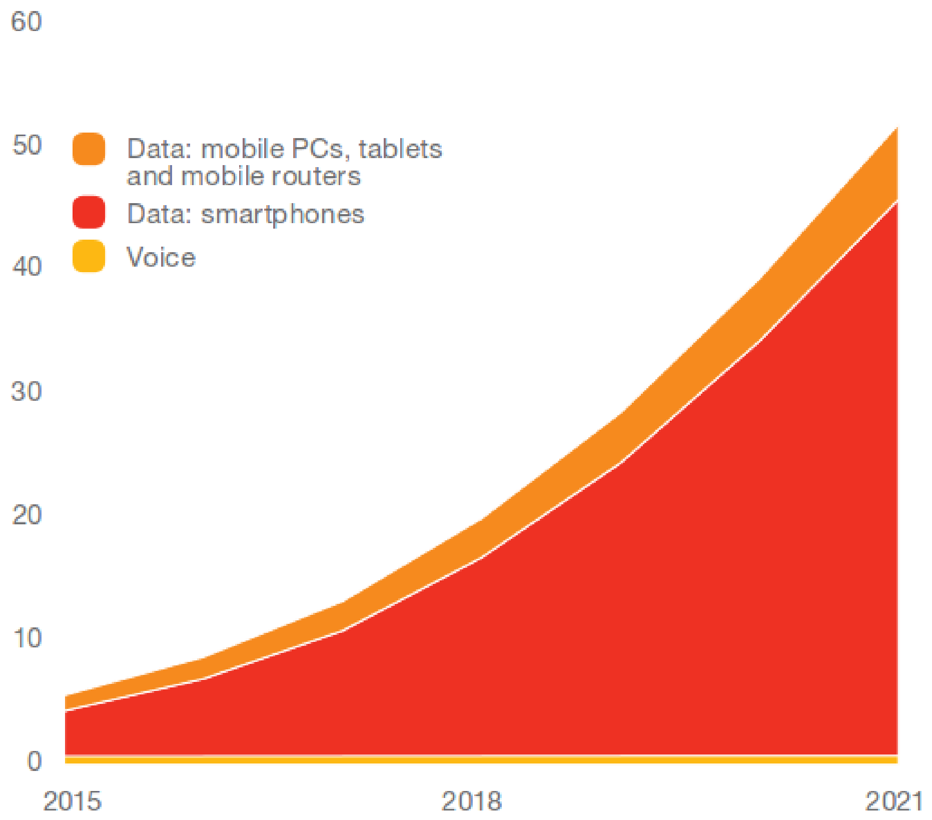


Figure 2.1: Mobile data traffic prediction 2015-2021 [1]

This forecast shows a significant increase in expected data consumption compared to previous reports: in 2014 the prediction for 2019 was around 20 ExaBytes/month [12] and now based on current network measurements the data consumption in 2019 will reach 30 ExaBytes/month. Again, this is due to the rise in subscriptions associated with Long Term Evolution (LTE) smartphones and the increasing data consumption per subscriber. Therefore network congestion caused by this rise in data demand is expected.

[2] www.youtube.com
[3] www.netflix.com
[4] www.hbo.com

2.1.2 QUALITY OF EXPERIENCE

Today's mobile clients want to be able to be "Always best connected" i.e want to connect to their operator's network and use services with good QoE. But as the number of smartphone users keeps rising, thousands of people gathered in "hot-spot" places such as airports, train and subway stations, stadiums, casinos or malls while using their devices specially during certain times of the day or days of the week lead to a significant increase in the cell network load causing congestion. Moreover, services ranging from voice or video calls or to share photos or even live-streaming video on social media[13] have different QoS requirements and during congestion, as most of the traffic (except Voice over LTE (VoLTE) and Voice over IP (VoIP)) is sent via a default bearer without guaranteed bit rate (as detailed in section 2.3) the user QoE gets affected as it is highly service or application dependent.

Poor service affects the user's opinion regarding their operator and cause a change to another service provider. In fact, according to a neuroscience study by Ericsson ConsumerLab [11], delays in loading web pages and videos lead to increased heart rates and stress levels comparable to watching a horror movie, for example, a single pause during video playback for re-buffering can increase stress levels by 15%. The study also concluded that when participants experienced no performance delays they increased the brand engagement with the service provider; otherwise when they faced medium delays they decreased the engagement with their service provider and favored competitors.

2.1.3 INFRASTRUCTURE COSTS / PLANNING

Globally operators invested heavily in their networks over these last few years, in 2014 the figure increased 9% to US$216 billion [4]. The push for 4G LTE networks implied a substantial investment in new Core Network and Radio Access Network (RAN) hardware but this new technology did not replace the need for 3G hardware as non-LTE devices exist in large numbers and new devices are multi-mode and support both technologies. Moreover, the development of 3G has not stopped with the introduction of LTE so operators must maintain and upgrade the legacy infrastructure to provide better service for non-LTE devices [14]. The adoption of LTE also mandated the acquisition of spectrum portions in the 800, 900, 1800, 2100 and 2600MHz bands in auctions with significant costs for operators [15] [16]. But it is relevant to note that the RAN backhaul of the network is where the overwhelming part of Capital Expenditures (CAPEX) spending is done as can be seen in Figure 2.2 [17].

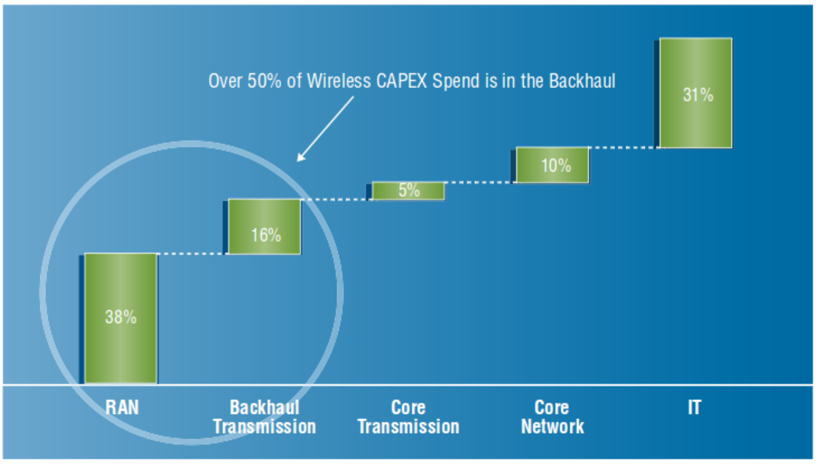


Figure 2.2: European MNOs CAPEX Breakdown [17]

The surge in data demand, specially in very crowded places, forced the deployment of extra macrocells but also overlays with small cells like picocells or femtocells [18] and Wi-Fi networks [19]. Moving forward to 2020 and towards 5G networks CAPEX is expected to keep growing globally while at a smaller rate as seen in Figure 2.3 totaling over US$1.4 trillion over this period [4].

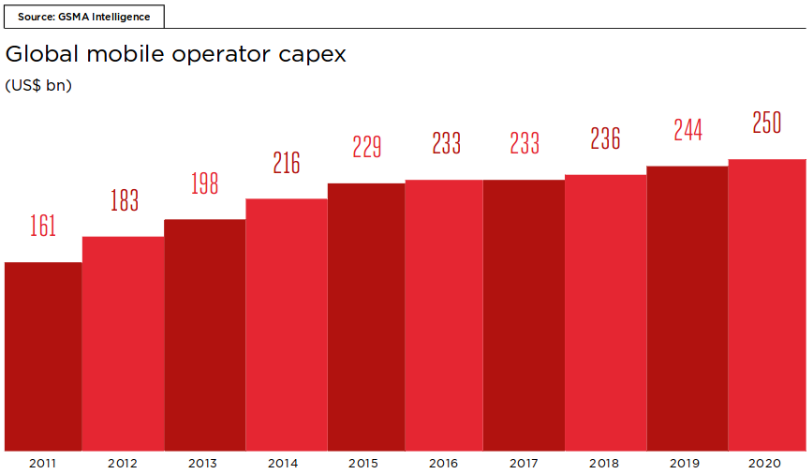


Figure 2.3: Global mobile operator capex (2011-2020) [4]

Revenue growth rate has been declining and operators are relying on the monetization of data traffic to invert the trend [4], so the dilemma of investing on infrastructure while keeping a healthy level of CAPEX to maintain profitability can only be solved by optimizing network resources.

Dimensioning the network for limited in time worst-case scenarios is not cost-effective because the required investment would result in an under-used network. It is therefore essential to manage the network congestion intelligently taking into account the subscriber profile (their billing plan) and the type of application and content.

2.2 STANDARDIZATION OF MOBILE NETWORKS

2.2.1 3GPP

The 3rd Generation Partnership Project[5] (3GPP) unites seven telecommunications standard development organizations (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC), with the goal of producing specifications regarding cellular telecommunications network technologies, including radio access, the core transport network, and service capabilities - including work on codecs, security, quality of service - therefore providing complete system specifications. The specifications also provide hooks for non-radio access to the core network, and for interworking with non-3GPP networks such as Wi-Fi or WiMAX. 3GPP has four main working Technical Specification Group (3GPP)s (TSGs) which are comprised of several Work Groups (3GPP)s (WGs) responsible for specific areas:

- Radio Access Networks (RAN) - responsible for the definition of the functions, requirements and interfaces of the UTRA/E-UTRA network in its two modes, Frequency Division Duplex (FDD) & Time Division Duplex (TDD);
- Service & Systems Aspects (SA) - responsible for the overall architecture and service capabilities of systems based on 3GPP specifications and, as such, has a responsibility for cross TSG co-ordination;
- Core Network & Terminals (CT) - responsible for specifying terminal interfaces (logical and physical), terminal capabilities (such as execution environments) and the Core network part of 3GPP system; and
- GSM EDGE Radio Access Networks (GERAN) - responsible for the specification of the Radio Access part of GSM/EDGE.

3GPP specifications and studies are contribution-driven by member companies such as vendors and network operators.

The SA WG2 Architecture (SA2) is responsible for Stage 2 of the 3GPP network architecture, it follows the services requirements (Stage 1) elaborated by SA WG1, and identifies the main functions and entities of the network, how they are linked to each other and the information they exchange. The output of SA2 is used as input by the groups who define the precise format of messages in Stage 3 (Stage 2 for the Radio Access Network is under TSG RAN's responsibility). SA2 has a system-wide scope and decides on how new functions integrate with the existing network entities.

The term "Stage" derives from the ITU's Telecommunication Standardization Sector (ITU-T)[6] method for categorizing specifications (Recommendation I.130 [20]).

- "Stage 1" refers to the service description from a service-user's point of view.
- "Stage 2" is a logical analysis, devising an abstract architecture of functional elements and the information flows amongst them across reference points between functional entities.
- "Stage 3" is the concrete implementation of the functionality and of the protocols appearing at physical interfaces between physical elements onto which the functional elements have been mapped.

[5] http://www.3gpp.org
[6] http://www.itu.int/en/ITU-T/Pages/default.aspx

3GPP uses a system of parallel "Releases", this allows the continuous development of current and new features. When a Release reaches the "freezing" stage no new features may be added. Release 13 which introduces among other features UPCON was frozen in March of 2016 by the SA2.

2.2.2 IEEE LAN/MAN 802.11 & the Wi-Fi Alliance

The Institute of Electrical and Electronics Engineers (IEEE) 802.11 standard is a set of specifications regarding the Media Access Control (MAC) layer and the Physical Layer (PHY) for implementing a Wireless Local Area Network (WLAN) communication in the 2.4, 3.6, 5 and 60 GHz frequency bands. The standardization work is performed by the IEEE LAN/MAN Standards Committee (IEEE 802)[7] WLAN Standards Working Group (802.11). Initially released in 1997, the standard has received subsequent amendments which are used as the basis for development of "Wi-Fi" products. These products are marketed according to the revisions of the standard as they introduce new features and capabilities.

Wi-Fi is a trademark term owned by the Wi-Fi Alliance[8] and refers to any WLAN product based on the IEEE 802.11 standards. Wi-Fi devices connect to a WLAN network, mainly using the 2.4 and 5 gigahertz unlicensed radio bands. The Wi-Fi Alliance, created in 1999, is a global non-profit industry association whose member companies sell Wi-Fi products. The Wi-Fi Alliance certifies these products for interoperability, industry-standard security protections and technology standards.

2.2.3 ITU-T

ITU[9] is the United Nations agency specialized in information and communication technologies – ICTs. ITU's Telecommunication Standardization Sector was created in 1865 and assembles experts from around the world to develop international standards known as ITU-T Recommendations which act as defining elements in the global infrastructure of ICTs and remains the world's only truly global ICT standards body.

ITU-T Recommendations (ITU-T Recs) - are standards defining how telecommunication networks operate and interwork. There are over 4000 Recommendations in force on topics from service definition to network architecture and security, from broadband DSL to Gbit/s optical transmission systems to next-generation networks (NGN) and IP-related issues.

ITU-T developed a framework of standards known as the International Mobile Telecommunication (IMT) system setting the specifications for global mobile communications. IMT-2000 referred to as 3G mobile technology was highly successful and widely deployed. ITU-T then launched the IMT-Advanced initiative as the 4G global mobile wireless broadband technology. After reviewing several proposals, ITU has determined that "LTE-Advanced" developed by 3rd Generation Partnership Project (3GPP) and "WirelessMAN-Advanced" developed by IEEE should be officially designated IMT-Advanced or true 4G technologies.

In brief, IMT-Advanced technologies follow these key points:

[7] http://www.ieee802.org/

[8] http://www.wi-fi.org

[9] http://www.itu.int/en/Pages/default.aspx

1. a high degree of commonality of functionality worldwide while retaining the flexibility to support a wide range of services and applications in a cost efficient manner;
2. compatibility of services within IMT and with fixed networks;
3. capability of interworking with other radio access systems;
4. high quality mobile services;
5. user equipment suitable for worldwide use;
6. user-friendly applications, services and equipment;
7. worldwide roaming capability; and,
8. enhanced peak data rates to support advanced services and applications (100 Mbit/s for high and 1 Gbit/s for low mobility were established as targets for research)

2.3 3GPP LTE NETWORK ARCHITECTURE

3GPP's Release'99 defined the basis for Universal Mobile Telecommunication System (UMTS) the 3rd Generation Mobile Network (3G) standards compliant with the IMT-2000 initiative.

In Release 8 3GPP introduced a new Network Architecture called the Evolved Packet System (EPS) which resulted from work to evolve both the radio access and the core network used in UMTS, but it should be noted that the work on UMTS evolution has not stopped and new features have been added in recent Releases to High Speed Packet Access (HSPA) and its evolution HSPA+ which data rates of 42.2 Mbit/s in the downlink or even higher with Advanced HSPA+ using multiple antennas in (Multiple Input Multiple Output (MIMO)).

LTE is the evolution of the Universal Terrestrial Radio Access Network (UTRAN) radio access therefore named Evolved UTRAN (E-UTRAN). The System Architecture Evolution (SAE) work evolved the non-radio aspects including the EPC network.

The "LTE" term is an European Telecommunications Standards Institute (ETSI)[10] trademark and is usually used to represent both the radio access and core network of the EPS architecture. In Release 10, the LTE-Advanced, and recently for Release 13, the LTE-Advanced Pro markers were introduced by 3GPP to recognize the significant enhancements in the architecture [21]. LTE-Advanced matched or surpassed the IMT-Advanced initiative requirements.

The EPS is essentially an all IP packet-switched network where real-time traffic such as voice and data are both carried in IP packets therefore eliminating the circuit-switched system used in UMTS for voice, moreover, the Access Network (the E-UTRAN) and the Core Network (the EPC) were functionally split into two entities managed by the RAN and SA 3GPP WGs and an overview of the network elements is shown in Figure 2.4.

[10] http://www.etsi.org/

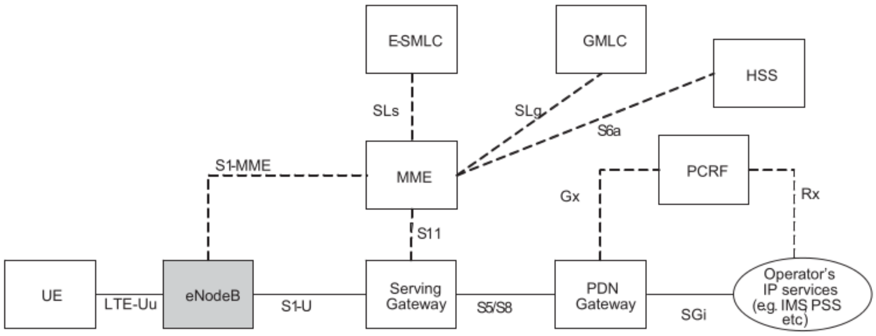


Figure 2.4: The EPS network elements [22]

Comparing with previous generations, Release 8 LTE brought new capabilities and key technologies [23]:

- Single-channel peak data rates of up to 300 Mbps in the downlink and 75 Mbps in the uplink;
- Improved spectral efficiency over legacy systems, particularly for the uplink;
- Full integration of Frequency Division Duplex and Time Division Duplex access modes;
- Packet-based Evolved Packet Core network to eliminate cost and complexity associated with legacy circuit-switched networks.
- Adoption of Orthogonal Frequency Division Multiple Access and Single Carrier Frequency Division Multiple Access for the downlink and uplink air interfaces to enable narrowband scheduling and efficient support of spatial multiplexing;
- Support for six channel bandwidths from 1.4 MHz to 20 MHz to enable high data rates and also efficient spectrum re-farming for narrowband legacy systems;
- Baseline support for Multiple Input Multiple Output spatial multiplexing of up to four layers on the downlink;
- Faster physical layer control mechanisms leading to lower latency.

The focus of this dissertation is at a network architecture level therefore the radio layer technologies used in LTE are not developed further. More detail can be found in [24] or [22]. The architecture Radio Access and Core Network elements are described in 3GPP TS 23.002 [25]; a summary of the relevant elements are presented in the following clauses.

2.3.1 Radio Access Network (E-UTRAN)

The Access Network in LTE is reduced to one node, the eNodeB (eNB) represented in Figure 2.5 that can responsible for managing multiple cells. The eNBs are normally inter-connected with each other by means of an interface known as X2, and to the MME by means of the S1 interface (S1-MME) and to the S-GW by means of the S1-U interface.

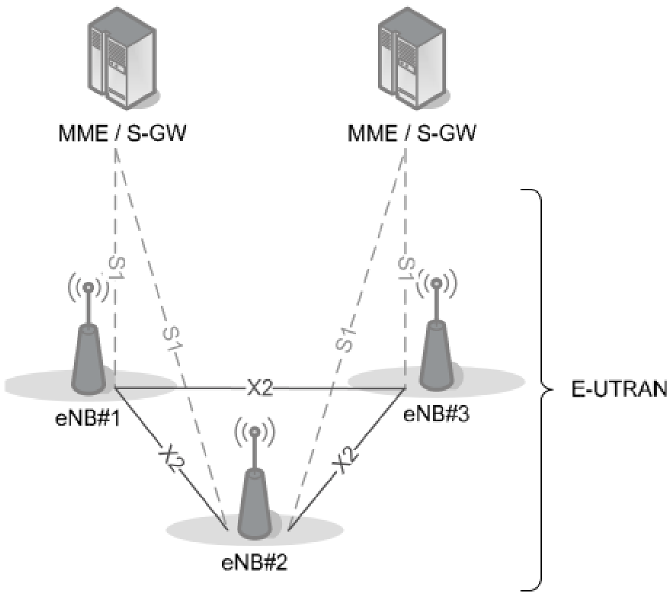


Figure 2.5: The E-UTRAN architecture [22]

The protocols which run between the eNodeBs and the UE are known as the Access Stratum protocols. On the network side, all of these functions reside in the eNBs. Unlike some of the previous second and third generation technologies, LTE integrates the radio controller function into the eNB.

The E-UTRAN (eNB) is responsible for all radio-related functions [22]:

- Radio Resource Management: All functions related to the radio bearers, such as radio bearer control, radio admission control, radio mobility control, scheduling and dynamic allocation of resources to UEs in both uplink and downlink.
- Header Compression: Ensure efficient use of the radio interface by compressing the IP packet headers which could otherwise represent a significant overhead, especially for small packets such as VoIP.
- Security: All data sent over the radio interface is encrypted.
- Positioning: The necessary measurements and other data are provided to the Evolved SMLC (E-SMLC) and assists the E-SMLC in finding the UE position.
- Connectivity to the EPC: The signaling towards the MME and the bearer path towards the SGW.

2.3.2 Core Network (EPC)

The EPC elements presented in Figure 2.4 are now described with their respective tasks in the network as in [22].

- The **Policy and Charging Rules Function (PCRF)** is responsible for policy control decision-making, as well as for controlling the flow-based charging functionalities in the Policy and Charging Enforcing Function which resides in the PDN Gateway. The PCRF provides the QoS authorization (QoS class identifier and bit rates) that decides how a certain data flow will be treated in the PCEF and ensures that this is in accordance with the user's subscription profile.

- The **Home Subscriber Server (HSS)** contains users' subscription data such as the EPS-subscribed QoS profile and any access restrictions for roaming. It also holds information about the PDNs to which the user can connect (APN or PDN Address). The HSS also holds dynamic information such as the identity of the MME to which the user is currently attached or registered.
- The **PDN Gateway (PGW)** is responsible for IP address allocation for the UE, as well as QoS enforcement and flow-based charging according to rules from the PCRF. The P-GW is responsible for the filtering of downlink user IP packets into the different QoS-based bearers. This is performed based on Traffic Flow Templates. The P-GW performs QoS enforcement for Guaranteed Bit Rate (GBR) bearers. It also serves as the mobility anchor for inter-working with non-3GPP technologies such as CDMA2000 and WiMAX networks.
- The **Serving Gateway (SGW)** serves as the local mobility anchor for the data bearers when the UE moves between eNodeBs. It also retains the information about the bearers when the UE is in idle state (known as EPS Connection Management IDLE (ECM-IDLE)) and temporarily buffers downlink data while the MME initiates paging of the UE to re-establish the bearers. In addition, the S-GW performs some administrative functions in the visited network, such as collecting information for charging (e.g. the volume of data sent to or received from the user) and legal interception. It also serves as the mobility anchor for inter-working with other 3GPP technologies such as GPRS and UMTS.
- The **Mobility Management Entity (MME)** is the control node which processes the signaling between the UE and the CN. The protocols running between the UE and the CN are known as the Non-Access Stratum protocols. The main functions supported by the MME are:

    1. Bearer management: the establishment, maintenance and release of the bearers, and is handled by the session management layer in the Non-Access Stratum (NAS) protocol.
    2. Connection management: the establishment of the connection and security between the network and UE, and is handled by the connection or mobility management layer in the NAS protocol layer.
    3. Inter-working with other networks: includes handing over of voice calls to legacy networks.

2.3.3 USER PLANE AND CONTROL PLANE

LTE's radio interface is characterized through its protocols where it can be defined by two main groupings according to the final purpose service: the user plane and the control plane protocols.

Even though separation of the control plane and the user plane was maybe one of the most important issues of LTE design, full independence of the layers is not feasible because, without interaction between the user plane and the control plane, operators are not able to control QoS, the source/destination of media traffic, and when the media starts and stops [26].

The User Plane carries user data through the Access Stratum (AS), all information sent and received by the UE, such as the coded voice in a voice call or the packets in an Internet connection, are transported via the user plane. User plane traffic is processed at different hierarchical levels, from eNB up to the core network (EPC). Also, control traffic is strictly tied to the user plane. This leads potentially to congestion which may impact user experience and therefore requires intelligent management:

For the transmission backbone it means the higher the level of network hierarchy the greater the amount of accumulated traffic generated. Therefore, higher level network elements will readily become the bottleneck of the network. Therefore, transmission capacity should be fitted to the network hierarchy; at higher levels high-capacity transmission means, such as fiber, are needed, but when it comes to the edge of the network microwave transmission becomes a more flexible and cost-effective substitution, particularly in terms of capacity extending. [26]

The Control Plane of the AS handles radio-specific functionalities. The AS interacts with the NAS, also referred to as the 'upper layers'. Among other functions, the NAS control protocols handle Public Land Mobile Network (PLMN) selection, tracking area update, paging, authentication and EPS bearer establishment, modification and release [22].

2.3.4 QoS in LTE

The EPS bearers shown in Figure 2.6 were introduced to implement QoS when routing IP traffic between a PDN Gateway and the UE. A bearer is in fact a virtual concept corresponding to the network configurations that provide priority treatment for some traffic e.g. VoLTE or VoIP packets over regular Internet traffic.

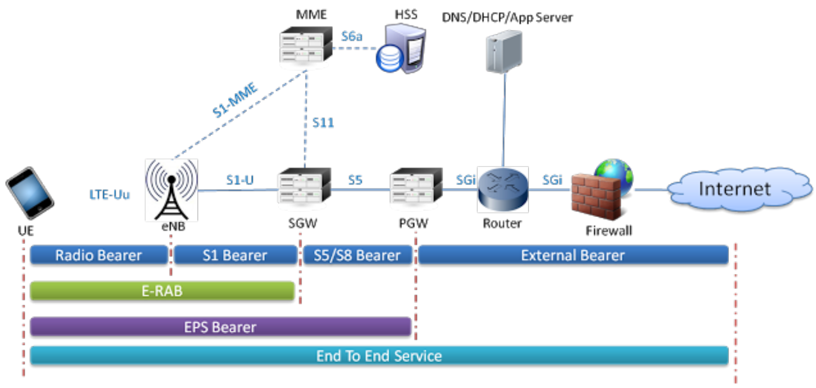


Figure 2.6: LTE EPS bearers [27]

There are two types of Bearer - the Dedicated bearer and the Default bearer with different properties as shown in Figure 2.7. At least one default bearer is established when a UE is attached to the LTE network and a dedicated bearer is established when there is need to provide QoS to a specific service (like VoIP, video etc).

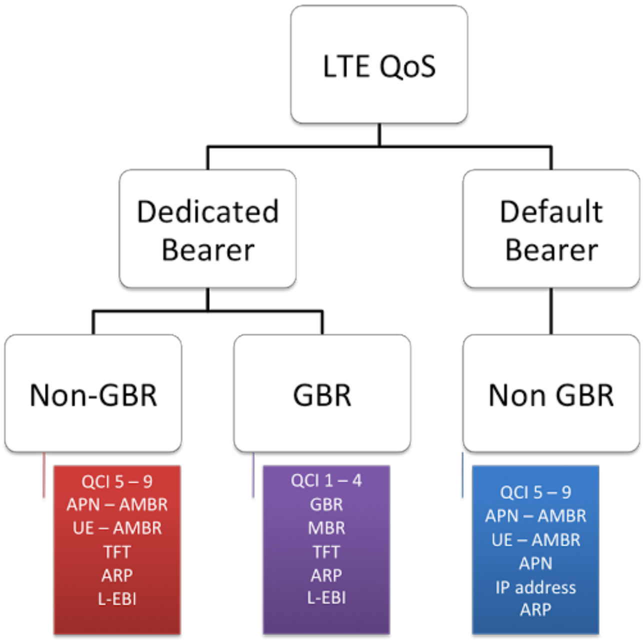


Figure 2.7: LTE EPS bearers [27]

The Dedicated bearer can be subdivided into Non-GBR and GBR types.
A GBR bearer provides guaranteed bit rate and is associated with parameters like GBR and MBR:

- GBR: The minimum guaranteed bit rate per EPS bearer. Specified independently for uplink and downlink.
- MBR: The maximum guaranteed bit rate per EPS bearer. Specified independently for uplink and downlink.

On the other hand, Non-GBR bearer does not provide guaranteed bit rate and has parameters like A-AMBR and UE-AMBR:

- A-AMBR: APN Aggregate maximum bit rate is the maximum allowed total non-GBR throughput to specific APN. It is specified interdependently for uplink an downlink.
- UE-AMBR: UE Aggregate maximum bit rate is the maximum allowed total non-GBR throughput among all APN to a specific UE.

Another parameter associated with all bearers is the QoS class of identifier (QCI) which defines IP level packet characteristics as shown in Table 2.1.

For services like VoLTE, where better user experience is required, dedicated bearers who use Traffic flow templates (TFT) to give special treatment to specific services are essential. Usually LTE networks with VoLTE implementations have two default and one dedicated bearer [27]:

- Default Bearer 1: Used for signaling messages (Session Initiation Protocol (SIP) signaling) related to IMS network. It uses QCI 5.
- Dedicated Bearer: Used for VoLTE VoIP traffic. It uses QCI 1 and is linked to Default Bearer 1
- Default Bearer 2: Used for all other smartphone traffic (video, chat, email, browser etc), assuming qci 9 is used here.

Table 2.1: Standardized QCI characteristics (adapted from Table 6.1.7 in [28])

QCI	Bearer Type	Priority Level	Packet Delay Budget	Packet Error Loss Rate	Example Services
1	GBR	2	100 ms	10^{-2}	Conversational Voice
2		4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3		3	50 ms		Real Time Gaming
4		5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)
65		0.7	75 ms	10^{-2}	Mission Critical user plane Push To Talk voice)
66		2	100 ms		Non-Mission-Critical user plane Push To Talk voice
5	Non-GBR	1	100 ms	10^{-6}	IMS Signaling
6		6	300 ms		Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p)
7		7	100 ms	10^{-3}	Voice,Video (Live Streaming) Interactive Gaming
8		8	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p)
9		9			
69		0.5	60 ms		Mission Critical delay sensitive signaling (e.g., MC-PTT signaling)
70		5.5	200 ms		Mission Critical Data (the same as QCI 6/8/9)

2.4 3GPP TRAFFIC MANAGEMENT MECHANISMS

3GPP has been introducing into its standard Releases several mechanisms intended to provide MNOs with traffic mobility solutions. These take advantage of new RAN technologies like femtocells or picocells and mobile devices with dual radios capable of having simultaneous 3GPP and WLAN connections. In brief:

- ANDSF - the mobile operator can send to the UE rules for discovering and selecting 3GPP or non-3GPP networks (Wi-Fi and WiMAX) to perform traffic offloading taking into account the devices' capabilities and user preferences.
- Local IP Access (LIPA) - introduced to provide access from UEs associated to a H(e)NB to other IP devices in the same residential/enterprise IP network without traversing the operator Core Network.
- Selected IP Traffic Offload (SIPTO) above RAN - provides the ability to offload internet traffic to a GW in the vicinity to the (H)eNB where the UE attaches while other traffic goes through the operator core network.
- SIPTO at the Local Network - introduced in two different architectures: one similar to LIPA and the other based on SIPTO above RAN. This function brought the internet traffic breakout to the residential/enterprise network.
- Multi Access PDN Connectivity (MAPCON) - allows the seamless mobility (with service granularity) of a PDN connection between the 3GPP and WLAN interfaces of a dual radio UE.
- IP Flow Mobility (IFOM) - the operator can offload traffic through the S2c interface to a WLAN access network exchanging seamlessly (with flow granularity) individual IP Flows over the WLAN and 3GPP interfaces of a dual radio UE.
- NBIFOM - enables IFOM mobility through the S2a and S2b interfaces with GTP and can be triggered by the UE or the core network of the operator.
- UPCON - the operator Core Network receives RAN User-plane congestion information which can be used by the PCRF serving the UE's PDN connections to perform policy decisions taking that congestion status into account in order to reduce the network congestion.

The focus of the 3GPP standardization effort has been on interworking with non-3GPP networks and soon WLAN networks will be further integrated in the network architecture.

2.4.1 Access Network and Discovery Function (ANDSF)

The ANDSF was introduced in 3GPP Release 8 and has been receiving updates to cooperate with RAN information and Hotspot 2.0 APs (Wi-Fi Passpoint[11]). The ANDSF function has the goal of assisting the User Equipment (User Equipment (UE)) in discovering and/or selecting non-3GPP networks like Wi-Fi and WiMAX favored by the mobile operator in roaming or non-roaming scenarios. ANDSF implements a Client-Server communication between the ANDSF Client (ANDSF-C) in the UE and the Server located in the operator Core Network through the S14 interface as represented in Figure 2.8.

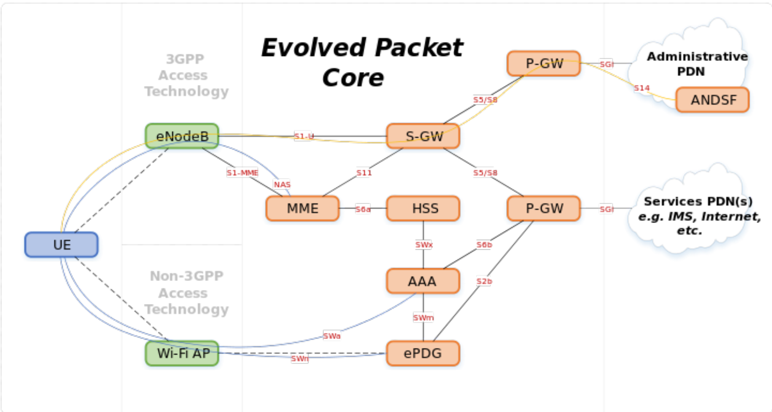


Figure 2.8: EPC with the ANDSF [29]

With ANDSF an operator is able to specify in detail how the UE shall prioritize, select and use the discovered Wi-Fi Hotspots or 3GPP networks. The types of policies may be based on the UE's location as well as other variables, such as time-of-day, subscription level, types of application etc. and are standardized in [30]:

1. Inter System Mobility Policy: used by devices that can connect over a single radio access at a given time to select the access type or network to connect to the EPC;
2. Access Network Discovery Information: a list of access networks available in the vicinity of the UE;
3. Inter System Routing Policy: used by devices to determine how to route specific IP traffic across multiple radio access networks;
4. Inter-APN Routing Policy: rules that determine which traffic should be routed through different PDN connections or non-seamlessly offloaded to WLAN;
5. WLAN Selection Policy: determine how a device should select and reselect a WLAN access network;

[11] http://www.wi-fi.org/discover-wi-fi/wi-fi-certified-passpoint

6. VPLMNs with preferred WLAN Selection Rules: contains a list of PLMNs that should be used by the UE when roaming;
7. Home Network Preferences: are provided the UE's home operator to assist the selection of a WLAN access network and a PLMN for 3GPP-based authentication over WLAN;
8. Visited Network Preferences: provided by another operator (not in the list of service providers equivalent to the home operator) when the UE is in roaming to assist the UE on selecting a PLMN for authentication via WLAN.

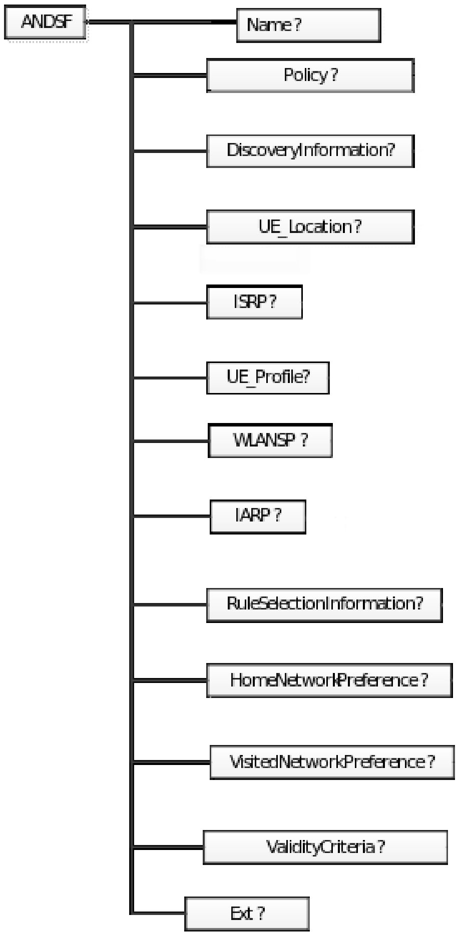


Figure 2.9: ANDSF Management Object [31]

Such policies are codified in the ANDSF Management Object (MO) shown in Figure 2.9, which is maintained in the UE and synchronized with the ANDSF Server (ANDSF-S) either in Push Mode when the Client initiates the procedure or in Pull Mode when the Server has the initiative. An example of an ANDSF policy tree is presented in Figure 2.10.

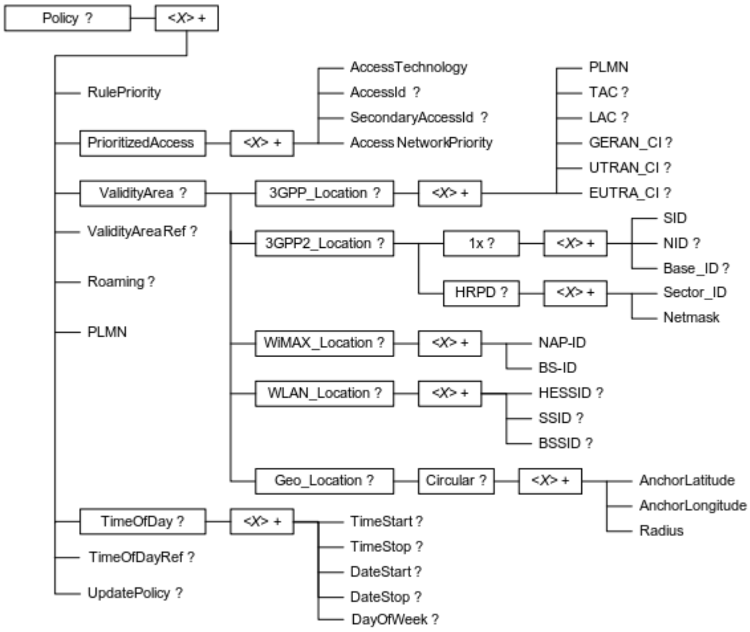


Figure 2.10: ANDSF Policy Management Object [31]

2.4.1.1 HOTSPOT 2.0

The Wi-Fi Alliance introduced the Hotspot 2.0 initiative with the goal of developing enhanced security, easier access, and more user-friendly operation for Wi-Fi products [32] certified as Wi-Fi Passpoint [33]. This technology uses a subset of the 802.11u interworking protocols. The 802.11u specification defines a number of enhancements to the 802.11 protocol to address the process of "interworking with external networks". With 802.11u a Wi-Fi client device is able to receive more information from the AP during the network scanning process as shown in Figure 2.11.

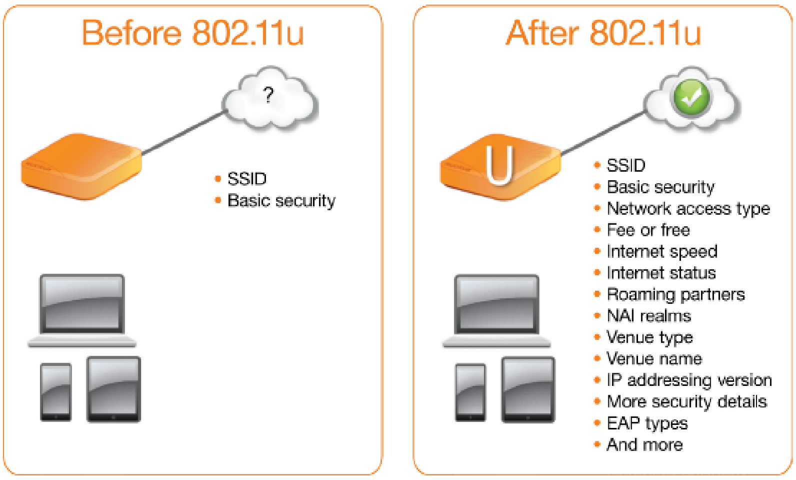


Figure 2.11: Network Discovery With and Without 802.11u [32]

Hotspot 2.0 enables an authentication into a Wi-Fi network via SIM/USIM therefore without the user intervention which is of great importance for mobile operators who want to offload traffic from their networks using ANDSF rules [34].

Two fundamental mechanisms are the basis for Hotspot 2.0: Generic Advertisement Service and Access Network Query Protocol(Figure 2.12).

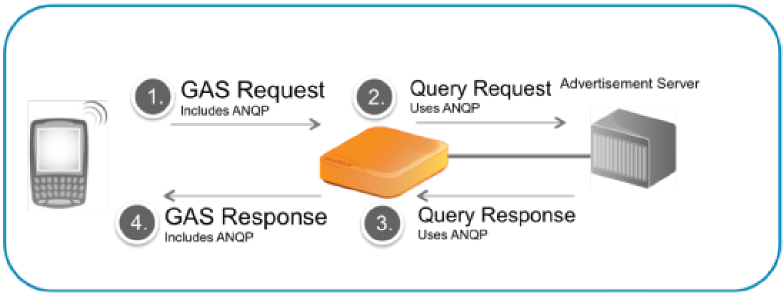


Figure 2.12: Logical Overview of the Hotspot 2.0 Advertisement Protocol Process [32]

ANQP is a query protocol specified in 802.11u used by stations to discover information about the network that is not advertised in beacons. GAS ia a framework that provides a frame exchange process (GAS Request/Response) and a framing format (using 802.11 Action frames) for ANQP advertisement services. One reason GAS is used is that prior to association, mobile devices have not obtained an IP address.

2.4.2 LOCAL IP ACCESS (LIPA)

The LIPA function was introduced into the 3GPP TS 23.401 [35] standard at Release 9 following the work in 3GPP TR 23.829 [36], with the goal of providing access for IP capable UEs, connected via a H(e)NB (e.g. indoor femtocell/picocell), to other IP capable entities/devices in the same residential IP network, without traversing traffic through the operator Core Network. It was enhanced in Release 10 to include enterprise networks.

This goal is achieved using a Local GW (L-GW) collocated with the HeNB as seen in Figure 2.13. The UE requests a new PDN connection to an APN for which LIPA is permitted, the network then selects the L-GW associated with the H(e)NB and enables a direct user plane path between them. Only control messaging/signaling is exchanged from the H(e)NB to the MME and, so far, no interface between the L-GW and the PCRF is specified. Dedicated bearers on the PDN connection used for LIPA are not supported and such request by the UE shall be rejected by the L-GW. LIPA may be provided by the VPLMN (the MME) when the UE is roaming only if the HSS indicates that LIPA roaming is allowed. If the UE's LIPA Closed Subscriber Group (CSG) authorization data for the APN has changed so that the PDN connection is no longer allowed in that cell, the MME must perform a release of the PDN connection to the APN.

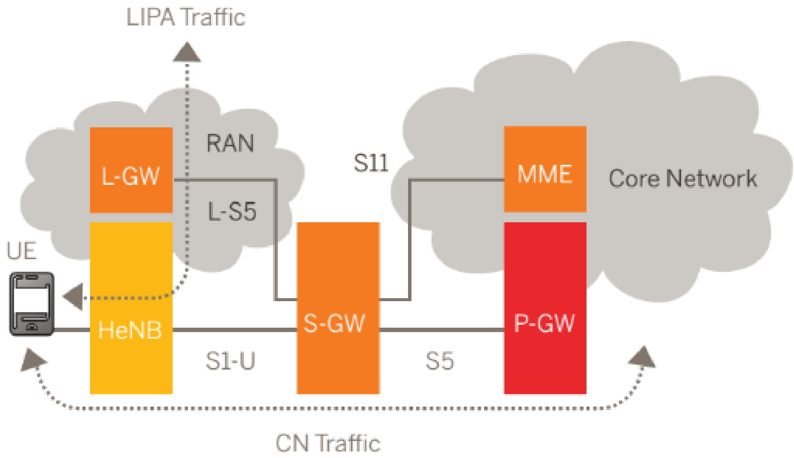


Figure 2.13: LIPA architecture with collocated L-GW [37]

Also, seamless mobility of the PDN connection between H(e)NBs is not supported because there is no possible anchor for the bearer used for LIPA during handovers. As such, the LIPA PDN connection is released by the L-GW when the UE moves away from the H(e)NB. After a trigger from the H(e)NB the L-GW initiates and completes the release of the LIPA PDN connection using the PDN GW initiated bearer deactivation procedure. After this procedure a handover to another H(e)NB may be initiated.

The collocated L-GW for this architecture does not support the Gx interface with the PCRF to receive dynamic Policy and Charging Control (PCC). However, it may support static QoS policies. Charging information in the EPS network is collected for each UE by the S-GWs and P-GWs, serving the UE. Charging for LIPA has to take into account that is has to compete with a free alternative like Wi-Fi so no charging or a flat rate based on subscription instead of session solution are preferable[37].

2.4.3 Selected IP traffic offload (SIPTO)

The SIPTO above RAN function was standardized in 3GPP TS 23.401[35] at Release 10 following the work in 3GPP TR 23.829[36] with the architecture in Figure 2.14 for the macro cell and in Figure 2.15 – for the Femtocell; then in Release 12, after the LIMONET study, SIPTO at the Local Network (SIPTO@LN) was introduced.

With SIPTO above RAN a mobile operator can offload selected types of IP traffic (best effort internet traffic) to a P-GW which is geographically/topologically close to the (H)eNB where the UE is attached while other traffic goes through the operator core network. The (H)eNB and the P-GW are located in the RAN of the mobile operator. The selection of the appropriate P-GW is therefore based on its proximity to the location of the UE.

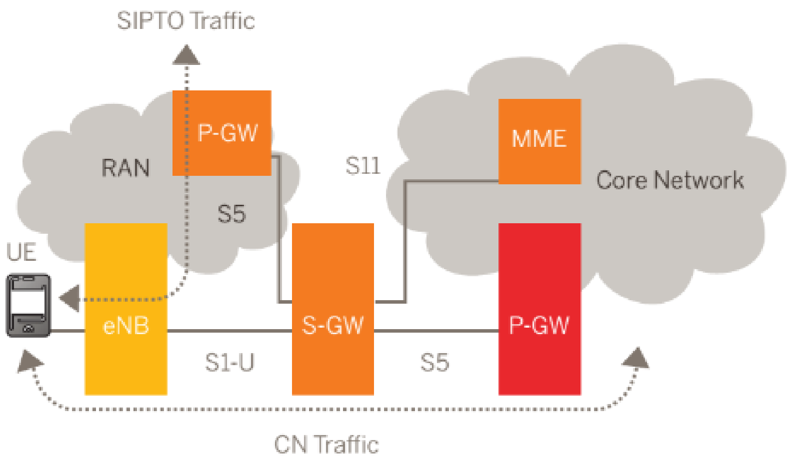


Figure 2.14: SIPTO above RAN architecture – macro network [37]

SIPTO is managed by the mobile operator on a per user and per APN basis by configuration of the subscription data in the HSS, which provides the MME with indication if the offload is allowed or prohibited. If the SIPTO permissions information from the HSS conflicts with MME's configuration for that UE, then SIPTO is not used. The MME may be configured on a per APN basis as to whether or not to use SIPTO (e.g. to handle the case where the HSS is not configured with SIPTO information for the UE).

If the MME detects UE mobility at a Tracking Area Update (TAU) procedure (detailed in 3GPP TS 23.401[35]) it may redirect a PDN connection towards a more appropriate GW, to do so it deactivates the PDN connection and indicates "reactivation requested" – this procedure causes disruption of services in case of a change of the IP address.

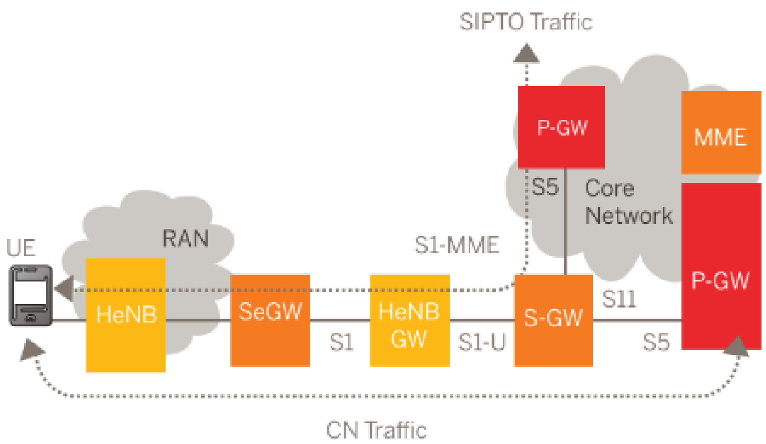


Figure 2.15: SIPTO above RAN with a Femtocell [37]

To address the issues of LIPA Mobility and SIPTO at the Local Network the LIMONET study was performed (3GPP TR 23.859[38]).

As conclusions of said study an architecture for LIPA mobility was proposed which introduced a 'Standalone logical L-GW' as a solution. Regarding the SIPTO at the Local Network – SIPTO@LN two different architectures were agreed. In the first a stand-alone GW (with S-GW and L-PGW collocated) is used and in the second the L-GW is collocated with the H(e)NB. At the standardization procedure for Release 12 the LIPA mobility solution was abandoned to focus on the agreed most important scenario: a per-APN SIPTO@LN to offload internet traffic without the requirement of IP address preservation in case of mobility. Both architectures for SIPTO@LN were introduced into 3GPP TS 23.401 [35].

With SIPTO@LN the traffic breakout point is now at the Local Network level reducing the load at the RAN. It enables an IP capable UE connected via a (H)eNB to access a defined IP network (e.g. the Internet) without the user plane traversing the mobile operator's network. The subscription data in the HSS is configured by the mobile operator per user and per APN to indicate to the MME if offload at the local network is allowed or not.

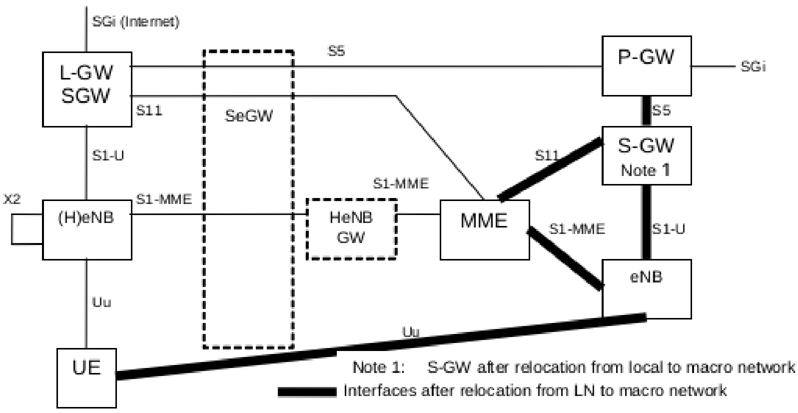


Figure 2.16: SIPTO@LN with a standalone GW (with L-PGW and SGW collocated) [35]

In the first standardized solution for SIPTO@LN in Figure 2.16, the stand-alone GW (with S-GW and L-PGW collocated) resides in the Local Network but it is assumed to be an "operator controlled service". The MME has the responsibility to select the optimal S-GW for the UE and perform the necessary relocation. For this solution the location of the Serving GW may be determined based on the operator policy and user's profile regarding support of SIPTO@LN so that at attachment to the (H)eNB, a local S-GW can always be selected independent of whether a PDN connection is established or not. If mobility is performed to the macro network without having a SIPTO connection, a S-GW relocation can be performed. At the attachment a macro S-GW may be allocated for PDN connection in the operator's network. If a new PDN connection is requested by the UE that requires that a local S-GW is selected to provide for SIPTO at the Local Network, S-GW relocation from the macro S-GW to the local S-GW shall be performed. IP data session continuity for SIPTO@LN is not supported so after the handover procedure the PDN must be disconnected by the MME. If charging data needs to be collected, the operator will have appropriate provisioning to support it in the SGW.

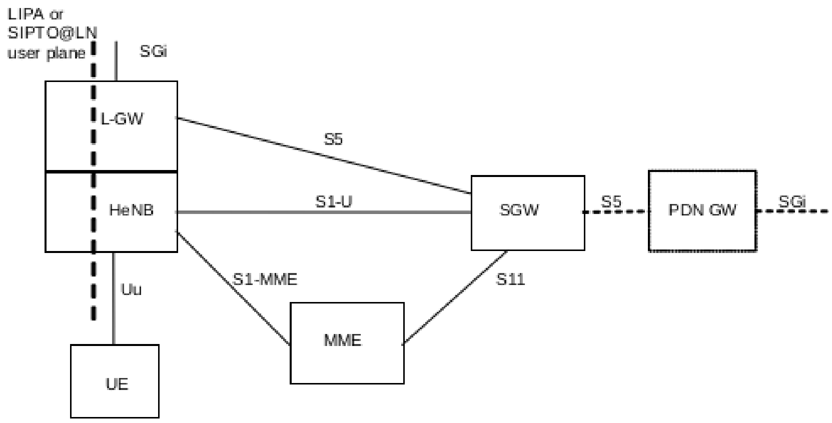


Figure 2.17: SIPTO@LN with L-GW collocated with the (H)eNB [35]

The second standardized solution in Figure 2.17, SIPTO@LN with L-GW function collocated with the (H)eNB reuses the LIPA architecture and as a mobility point of view it suffers from the same issues as LIPA – as the UE moves away from the HeNB and approaches another HeNB the PDN connection must be dropped and be re-established without IP preservation. The L-GW information for SIPTO@LN is signalled on S1 separately from the L-GW information for LIPA. The L-GW shall be able to discriminate between PDN connection for SIPTO at the Local Network and for LIPA.

2.4.4 IP FLOW MOBILITY (IFOM)

IFOM was introduced in 3GPP Release 10 as a result of the study for the support of IP flow mobility for the S2c interface performed in TR 23.861 [39] which intended to develop a mechanism so that a mobile operator could offload certain traffic to a non-3GPP access network with service continuity. The results of this work were transferred to TS 23.261 [40] which provides the system description for IP Flow Mobility between a 3GPP network and a WLAN network.

Non-3GPP networks may be 'trusted' if they connect to the EPC directly to the PDN-GW or 'non-trusted' if they connect to the PDN-GW through the ePDG (additional security and authentication is required).

The S2c interface as seen in Figure 2.18 is used by the UE to connect to the P-GW for both trusted and non-trusted networks.

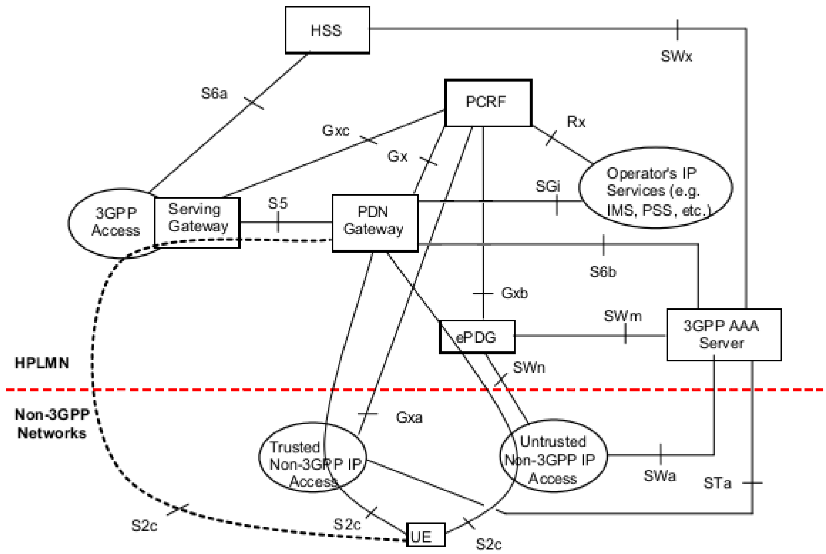


Figure 2.18: Architecture within EPS using S5, S2c [39]

Unlike LIPA or SIPTO, IFOM requires a dual radio UE capable of connecting simultaneously to the same PDN connection over its 3GPP and Wi-Fi interfaces. IFOM is implemented based on the DSMIPv6 protocol[41] over the S2c interface therefore it allows the exchange of different (single or multiple) IP flows between the two accesses with no service break (with IP address preservation and session continuity). This gives the mobile operator the ability to perform WLAN offload - determine which IP flows may be routed to a WLAN network (usually best effort traffic such as internet) and which must remain in the operator network (traffic with higher QoS requirements so congestion in radio and core network are improved. The flow mobility (adding, deleting or moving flows between accesses) is performed by the UE so the mobile operator must provide to the UE routing rules. These rules may be provided statically but also “dynamically” (in Push of Pull modes) via ISRP (or IARP) rules using the ANDSF function.

2.4.5 MULTIPLE ACCESS PDN CONNECTIVITY (MAPCON)

The MAPCON function also requires a UE with dual radios connected to WLAN and 3GPP simultaneously but unlike IFOM only entire PDN connections to various APNs can be routed, i.e. a PDN connection to an Internet APN can be routed through a 3GPP access or offloaded through an (untrusted) Wi-Fi network. MAPCON is mainly used to offload traffic from the core network. Mobility sensitive applications (e.g. VoIP, Video streaming) shall not be offloaded as IP connection may fail during handover.

2.4.6 NETWORK-BASED FLOW MOBILITY (NBIFOM)

A major disadvantage of IFOM is that it relies on the UE to provide the flow routing policy as part of the mobility signaling to the Home Agent (the PDN gateway) and when a mobile operator wants to change the routing rules it must do so through the ANDSF with ISRP or IARP rules (2.4.1). As the ANDSF currently has no (standardized) interface to the Policy and Charging Control (PCC) (TS

23.203 [28]) the updated flow policies must be provided to the network by the UE and as the network context and resource availability may have meanwhile changed the PCRF will not be able to authorize the new policies. TR 23.861 [39] was revived to study the implementation of NBIFOM for the S2a and S2b interfaces in Release 13. The work on the TR was presented for plenary approval in July of 2015 (SP-68) and served as the basis for 3GPP TS 23.161 the Stage 2 specification for NBIFOM approved later in September's plenary meeting for Release 13 and is currently in version 13.2.0 [42].

For NBIFOM, the UE and the PDN GW exchange Routing Rules (RR). Routing Rules (RR) describe the mapping between a set of IP flows (identified by a Routing filter) and a Routing access type (e.g. 3GPP or WLAN) over which to route these IP flows.

Routing rules are exchanged over control plane protocols (NAS over a 3GPP access, WLCP / IKE over WLAN access, S2a/S2b/S5/S8 control protocol). Rule installation is possible at any time during the lifetime of a PDN connection or even if there is no IP flow currently matching the routing filter in the Routing Rule.

The following NBIFOM modes are supported:

- **UE-initiated NBIFOM**: In the UE-initiated NBIFOM mode only the UE controls the traffic routing within the multi-access PDN connection. The UE determines the Routing rules and provides them to the PDN GW. The network may either accept or reject the Routing Rules requested by the UE, but does not provide Routing rules itself.
- **Network-initiated NBIFOM**: In the Network-initiated NBIFOM mode, the network controls the traffic routing within the multi-access PDN connection The PDN GW determines the Routing rules and provides them to the UE. The UE may either accept or reject the network's request for Routing rules (e.g. based on the suitability of the WLAN link conditions), but does not provide Routing Rules itself.

The mode applied to a PDN connection is determined during the PDN establishment procedure.

2.5 CHAPTER CONSIDERATIONS

This chapter presented the state of the art on congestion management in mobile networks starting with its drivers and the standardization entities. The focus was placed on 3GPP's technologies namely LTE and its network architecture. Then several traffic mobility and offloading 3GPP mechanisms were detailed. It was noted that the User Plane of the LTE architecture is the bottleneck as it serves all the traffic sent from and to the UE. Congestion awareness and management was the target of 3GPP's work for the standardization of UPCON which will be detailed in Chapter 3. The ANDSF is of great importance to operators who manage Wi-Fi networks as it enables them to offload traffic from their cellular network. It would be of great importance to evaluate its use in congestion management.

CHAPTER 3

3GPP User Plane Congestion Management

This chapter presents in detail an overview of the result of 3GPP's standardization work on the UPCON function for Release 13.

3.1 Standardization Work

3GPP performed a feasibility study on user plane congestion management [43] to address several scenarios or use cases where user plane congestion occurs and propose the management requirements. The agreed upon requirements where then introduced into the Service aspects and Service principles technical specification (3GPP TS 22.101 [44])

According to the study, RAN user plane congestion occurs when the demand for RAN resources to transfer user data exceeds the available RAN capacity to deliver the user data for a significant period of time, in the order of few seconds or longer. Therefore, a short-duration burst of traffic is not considered as such. The user plane congestion may be due to full use of cell capacity, as in Figure 3.1, or due to 3GPP RAN to EPC interface capacity limitation as in Figure 3.2. It is up to the operator to define the level of RAN resources utilization that indicates that the RAN is user plane congested.

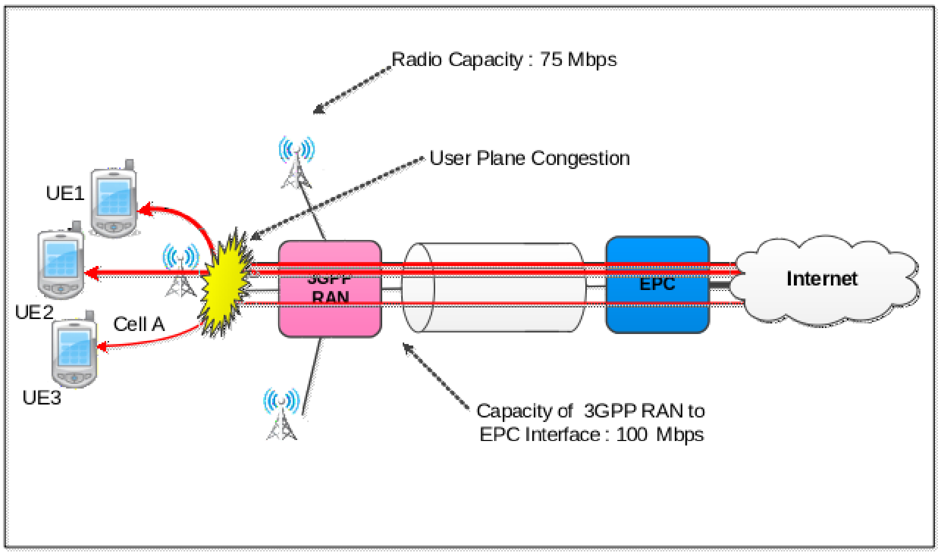


Figure 3.1: User plane congestion due to full use of cell capacity (example capacities) [43]

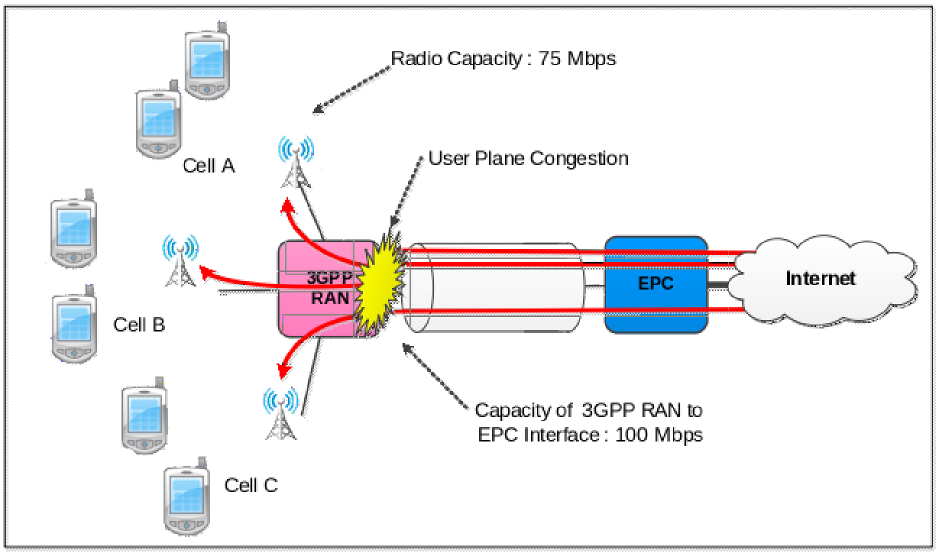


Figure 3.2: User plane congestion due to 3GPP RAN to EPC interface capacity limitation (example capacities) [43]

According to the standard[44], the user plane congestion management (mitigation) must be performed:

- Prioritizing traffic,
- Reducing traffic, or
- Limiting traffic.

Traffic shall be prioritized according to the QoS requirements of different applications (e.g., social networking, video, games, software updates, etc.), communications with preferential treatment and the type of user (e.g., heavy user, roaming user, etc.).

To reduce traffic load, the network can compress data of specific applications or adapt real time communications so that they consume less bandwidth (voice and video can be treated differently) taking into account the user information, so that even with congestion a high profile user (e.g. “platinum” subscription) must have a good experience.

The network may also handle traffic in order to prohibit or delay all or a particular traffic, depending on whether a service request is for Unattended Data Traffic or Attended Data Traffic, or limit traffic from operator-controlled and/or third party services such as Push services.

A follow up study on system enhancements for user plane congestion management (3GPP TR 23.705[45]) was performed to introduce UPCON in Release 12, but work was abandoned at the SP-62 meeting for lack of agreement. The TR was later resuscitated by SA2 in a March 2014 meeting and UPCON was scheduled for Release 13. The final version of this report was presented in December of 2014.

The following three Building Blocks (BB) with different tasks were envisioned to be worked in parallel but work on BB2 and BB3 was stopped in the SA2 SP-64 plenary meeting until the completion of BB1 for time budget issues[46]:

- Building Block 1 - RAN Downlink Traffic Differentiation, Congestion Detection and Reporting (UPCON-DOTCON),
- Building Block 2 - Application and Content Prioritization (UPCON-CON) and
- Building Block 3 - Uplink and attended/unattended traffic (UPCON-TRAF).

The report was then revised to include only Building Block 1 which addressed RAN User plane Congestion awareness and mitigation as well as differentiated treatment for non-deducible service data flows working on these features:

- RAN Congestion Detection: Determine if the RAN is congested. This includes considerations of ’what’ congestion is, ’when’ (to distinguish between transient and sufficiently long periods to take action, e.g. to report congestion).
- RAN Traffic Differentiation: This includes enhancements to traffic differentiation (e.g. marking, priority, etc.) and considerations for shared networks.
- RAN Congestion Reporting: Determines how (which method) and what will be reported from the RAN to the Core Network.
- Core Network (CN) Congestion Mitigation mechanisms.
- Dynamic Policy Control: Provide policies for RAN Traffic Differentiation and CN Congestion Mitigation, per subscriber policies, including Application/Rx handling.

The solutions presented in the report were Core Network based (CN-based), RAN-based, a combination of both and UE-based but ultimately the CN-based solution with off-path based RAN user plane congestion reporting shown in Figure 3.3 and mitigation based on PCRF policy decisions were chosen for standardization.

In brief, with this solution the RAN Congestion Awareness Function (RCAF) is introduced to the PCC logical architecture to receive RAN User Plane Congestion Information (RUCI) from the RAN

OAM and the MME or SGSN serving the UEs and provide that information to the PCRF to be used in congestion management policies.

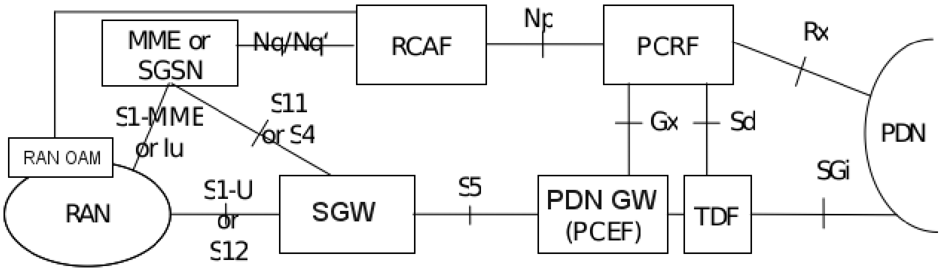


Figure 3.3: EPS Network architecture with UPCON off-path RUCI reporting [43]

This solution clearly has the merit of preventing the overflow in the MME-SGW-PGW-PCRF control path caused by the congestion signaling.

3.2 EPC ARCHITECTURE UPGRADE

The RAN Congestion Awareness Function is a new functional element in the Release 13 Policy and Charging Control logical architecture, shown in Figure 3.4 for a non-roaming scenario, that interfaces with the PCRF via the Np interface as standardized in 3GPP TS 29.217[47], with the MME via the Nq interface and with the SGSN with the Nq' interface specified in 3GPP TS 29.405[48] and with the RAN OAM via a non-standardized interface to gather RAN user-plane congestion information.

The RCAF has the goal of collecting RAN user-plane congestion information from the RAN OAM and the MME(S) or SGSN(s) serving the congested area and packaging it as RUCI and provide them to the PCRF with the Np protocol. This information allows the PCRF serving the UE's PDN connections to perform policy decisions taking that congestion status into account in order to reduce the network congestion. The PCRF acts only on the operator Core Network elements, i.e. congestion mitigation policies are not applied at the RAN level or in the UE.

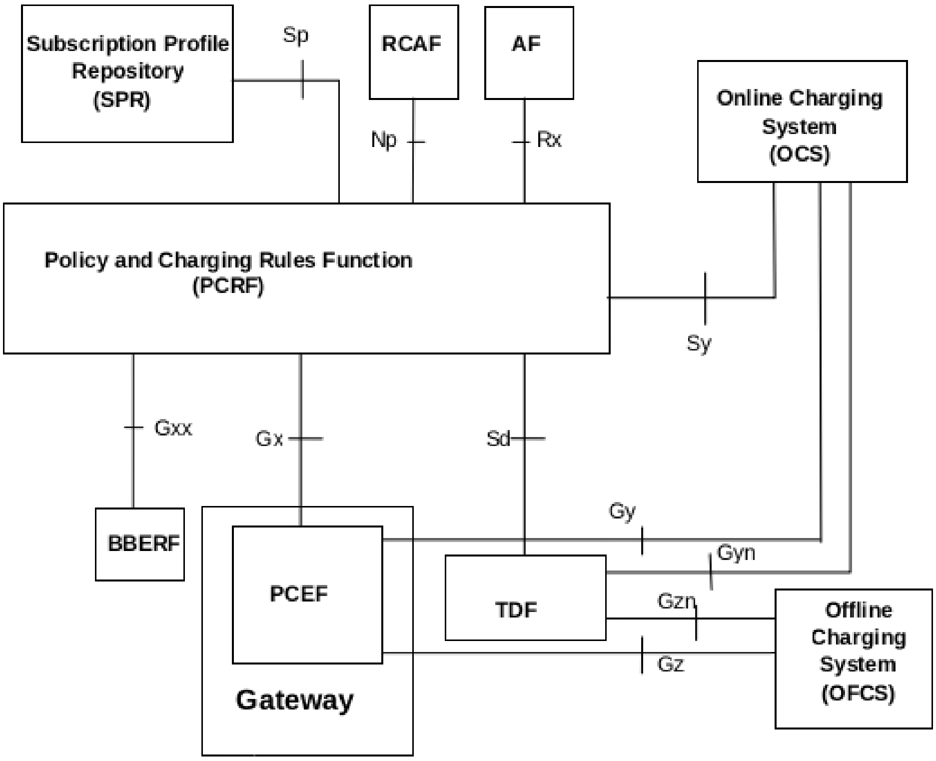


Figure 3.4: PCC Release 13 Architecture [28]

The Traffic Detection Function (TDF) and the Sd interface were introduced in Release 11 to enable application detection, gating, QoS and application based charging and if deployed are used for congestion mitigation.

3.2.1 CONGESTION REPORTING

The RUCI includes the IMSI identifying the UE impacted by congestion and the eNB identifier, ECGI or SAI identifying the eNB, E-UTRAN cell or Service Area, respectively, serving the EU when location reporting is enabled, as well as the PDN ID for which congestion information is reported and the corresponding congestion level or a "no congestion" state indication.

A single logical PCRF entity may be deployed by means of multiple and separately addressable PCRFs. In networks with multiple PCRFs a Diameter Routing Agent (DRA) is typically deployed and the RCAF uses the DRA to contact the PCRF(s) as follows:

- When the Gx gets established, the DRA "assigns" a PCRF for a given IMSI/APN combination and remembers the relationship (IMSI, APN and selected PCRF).
- Based on this relationship DRAs support finding the PCRF serving an IMSI/APN combination.

This is based on the functionality described in 3GPP TS 23.203 [28] or 3GPP TS 29.213[49].

Two types of RUCI messages are used to send congestion information from the RCAF to the PCRF serving that user ID and PDN ID:

- Non-aggregated RUCI report messages, which are sent on a per-UE and per-APN basis using DRA routing. The IMSI and the APN are used to route messages to the serving PCRF.

- Aggregated RUCI report messages, which contain congestion information for different user IDs and PDN IDs served by the same PCRF. The serving PCRF ID determines the destination of the message.

The RCAF must send a Non-aggregated report when it does not know the destination PCRF for a specific user ID and PDN ID, otherwise either type of report may be used.

The 3GPP specification does not define which information should be contained in a single aggregated RUCI report message. In this way, the RCAF may, for example, aggregate into a single message information only for a given cell or eNB, or the RCAF may wait for a configurable period of time to aggregate information from multiple cells or eNBs into a single message.

The RCAF shall report the RUCI to the PCRF according to the configured interval when at least one of the following conditions applies, unless the RUCI reporting is disabled for the PDN ID or for the user ID and the PDN ID:

- The RCAF detects a UE in the congestion area for the first time.
- A reporting restriction is enabled and the congestion level set ID is changed.
- A reporting restriction is not enabled and the congestion level value is changed.
- A conditional restriction to restrict location reporting is not enabled and the UE is in a congested area and the location is changed.
- The RCAF detects that the UE is no longer experiencing congestion (i.e. the UE is no longer detected in any of the congested cells that the RCAF is monitoring).

The RCAF maintains a context per UE and APN which is identified by the IMSI and the APN with the following information:

- The previously reported congestion level.
- The reporting restrictions received from the PCRF. The reporting restrictions are stored by the RCAF until the PCRF explicitly signals to remove the reporting restrictions.
- The logical PCRF ID (The PCRF id of the PCRF that serves the user ID/PDN ID when multiple PCRFs are deployed) received from the PCRF to identify the PCRF that is the Np destination used to send aggregated reports.

Depending on the reporting interval, a UE can move outside the area indicated in the report without the RCAF immediately notifying the PCRF. Therefore, the PCRF may receive conflicting information about the cell serving a UE via Np and Gx interfaces, when the location change reporting is enabled for that UE. In this situation, the information received via Gx shall take precedence. The PCRF can disable location reporting to prevent this case.

According to the specification, the “ReportRestriction” is an optional feature but, if both the PCRF and the RCAF support it, the PCRF has the task of adding, updating and removing reporting restrictions on the RCAF. This is done through the Np interface with specific Diameter commands on a per-UE and per-APN basis. These restrictions may be used to limit or disable the reporting to certain APNs, or both user ID and PDN ID, the congestion location, or to define sets of non-overlapping congestion levels, such that the report indicates the congestion is within a given set instead of the level itself in order to reduce the necessary signaling.

3.2.2 CONGESTION MITIGATION

Once the PCRF is provisioned with RUCI by the RCAF via the Np interface it may take that information into account when sending congestion management policies to other Core Network functions as shown in Figure 3.5.

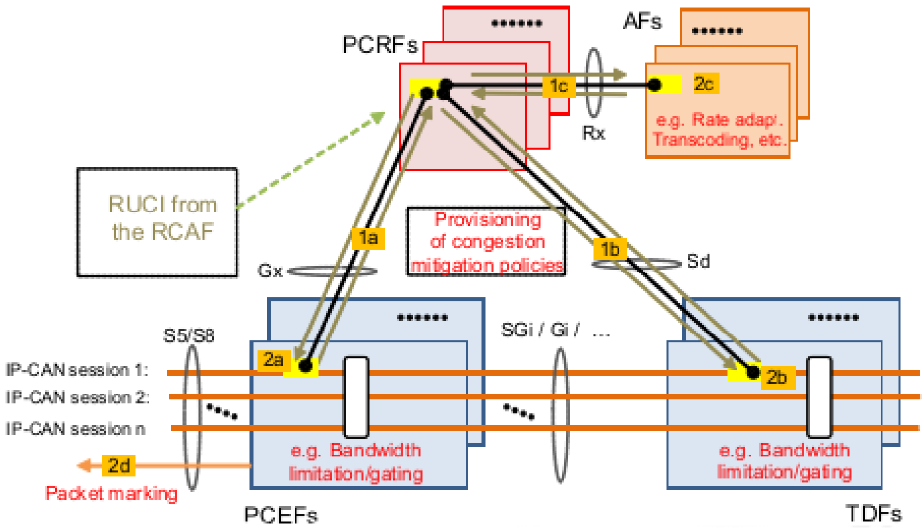


Figure 3.5: Provision of congestion mitigation policies [43]

Policies are sent to the PCEF over Gx, to the TDF over Sd or to the AF over Rx. These policies can be provisioned before the congestion occurs or after the PCRF becomes aware of the congestion level. All the existing variants of policy provisioning (predefined and activated/de-activated dynamically and provided dynamically) may be used for congestion mitigation.

Additionally, for the Rx interface between the PCRF and the AF, a Retry-interval AVP (AVP code 541 is of type Unsigned32) is introduced which includes a time interval in seconds so that, according to the congestion level, the PCRF can send to the AF the ‘Retry interval’, via an Rx Diameter session, which indicates a time interval in seconds that the AF must wait until it retries to send the same service information to the PCRF (for the same IP-CAN session) when the service information is temporarily rejected by the PCRF. According to 3GPP TR 23.705 clause 6.1.6 [45], this interval is calculated based on heuristics and is implementation dependent: it cannot accurately predict the length of the congestion but it prevents the immediate increase in congestion.

RAN user plane congestion mitigation also uses existing bandwidth limitation parameters. The PCRF sends to the AF in an AA-Answer command during the Rx session establishment the Max-Requested-Bandwidth-DL AVP, Max-Requested-Bandwidth-UL AVP within the Acceptable-Service-Info AVP.

3.2.3 UPCON SIGNALING FLOW

The signaling flow for UPCON reporting is presented in Figure 3.6 and described below. The messages exchanged between the RCAF and the PCRF and between the RCAF and the MME or SGSN are presented in detail in 3.4.1 and 3.4.2 respectively.

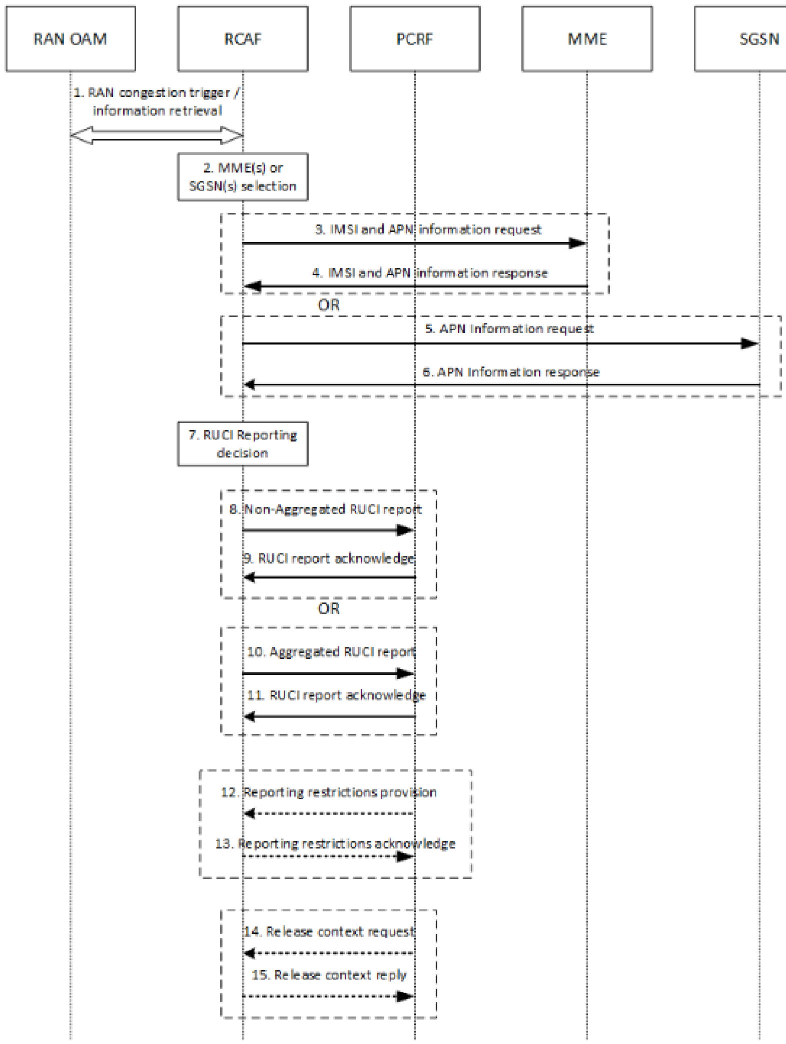


Figure 3.6: Network architecture with UPCON off-path RUCI reporting

1. Via a non-standardized interface, the RCAF retrieves congestion related data from the RAN OAM managing the geographical area it serves, such as cell load as well as the affected area location (e.g. eNodeB ID, E-UTRAN cell ID or UTRAN Service Area ID). For UTRAN, the RCAF also receives the list of UEs (IMSIs) impacted by a change of RUCI in a cell from the RAN's OAM provided by the RAN (the IMSI is sent by the SGSN over Iu in RANAP Common Id message).

2. With the received congested area location information from the RAN OAM,
   - the RCAF selects the MME(s) serving the eNodeB or E-UTRAN cell affected by user plane congestion based on their Tracking Area Identity (TAI) through the DNS procedure specified in sub-clause 5.4 of 3GPP TS 29.303 [50].
   - the RCAF selects the SGSN(s) serving the user plane congested UTRAN Service Area through the DNS procedure specified in sub-clause 5.5.2 of 3GPP TS 29.303 [50].

3. In case of E-UTRAN, the RCAF sends an **NqAP-IMSI-APN-INFORMATION-REQUEST** message to the MME via Nq, indicating the eNodeB ID or ECGI Cell ID (according to operator configuration) for information on the UEs and APNs affected by RAN user plane congestion.

4. The MME replies with an **NqAP-IMSI-APN-INFORMATION-RESPONSE** message containing the IMSIs of the UEs which are currently in ECM-CONNECTED state and having active E-RABs under a given eNodeB or E-UTRAN cell and for each IMSI the list of APNs currently having active PDN connections.
5. In case of UTRAN, the RCAF sends an **NqAP-APN-INFORMATION-REQUEST** message to the SGSN via Nq' for information on the UEs affected by RAN user plane congestion information providing their IMSIs.
6. The SGSN replies with an **NqAP-APN-INFORMATION-RESPONSE** message with the APNs that are currently having PDP contexts activated for each IMSI provided by the RCAF.
7. Based on the IMSI/APN information received from the MME(s) or SGSN(s), the RCAF shall send Non-Aggregated or Aggregated RUCI reports if:

   - At least one of the reporting conditions applies.
   - RUCI reporting is not disabled for at least one of the PDN IDs or user IDs and the PDN IDs that met those conditions.

   To send RUCI regarding a single UE (user ID and PDN ID) the RCAF must send a Non-aggregated RUCI report if it does not know the destination PCRF.

   The RCAF shall send an Aggregated RUCI report to send RUCI of different UEs (user IDs and PDN IDs) that it knows share the same destination PCRF.
8. The RCAF sends a non-aggregated RUCI report to the PCRF via Np with a Diameter Non-Aggregated RUCI Report Request (NRR) command. The NRR command includes the user ID, the PDN ID, a congestion level set ID or congestion value if a reporting restriction was provided earlier or not, the congestion location if this report is enabled and the RCAF ID. The RCAF stores or updates the UE context.
9. The PCRF stores the information and sends as an acknowledgment a Diameter Non-Aggregated RUCI Report Answer (NRA) command including the PCRF id to the RCAF. The PCRF may also provide reporting restrictions for that user ID in the NRA command if the "ReportRestriction" feature is supported. The RCAF updates the UE context to store the PCRF id and reporting restrictions.
10. The RCAF sends via Np an Aggregated RUCI report with a Diameter Aggregated RUCI Report Request (ARR) command. The ARR command includes the PCRF ID, one or more Aggregated-RUCI-Report AVP with a congestion level set ID or congestion value if a reporting restriction was provided earlier or not (each AVP aggregates the users that share the same level of congestion or share the same congestion level set ID), the PDN ID and the aggregated congestion information. The RCAF stores or updates the UE context for all the UEs listed in the report.
11. The PCRF stores the information and sends as an acknowledgment a Diameter Aggregated RUCI Report Answer (ARA) command back to the RCAF.
12. If the "ReportRestriction" feature is supported, the PCRF can at any time specify, modify or disable restrictions for RUCI reporting for a specific user ID and PDN ID sending a Diameter Modify Uecontext Request (MUR) command to the RCAF.

13. The RCAF acknowledges the reporting restriction provisions with a Diameter Modify Uecontext Answer (MUA) command to the PCRF and changes the UEs and APNs context.

14. If the “ReportRestriction” feature is supported and the PCRF receives a non-aggregated RUCI report for the same user ID and PDN ID but with a different RCAF id than the previous report it must update the congestion information and after the NRA reply perform the removal of the UE context from the old RCAF sending a Diameter MUR command with “Release Context” to that RCAF for that user ID and PDN ID.

15. The RCAF responds with a Diameter MUA command and releases the context corresponding to the given user ID and PDN ID including any reporting restriction.

When the PCRF performs a Removal of UE context procedure for a specific user ID and PDN ID and sends a MUR command to the RCAF with a “Release Context” indication, RUCI exchanges may be ongoing so in this case the request shall be handled immediately and the context released. As for the PCRF, if a MUR command is being handled and a NRR command is received for that specific user ID and PDN ID then this request shall be rejected with a Diameter experimental result code of DIAMETER_PENDING_TRANSACTION. If a MUR command is being processed at the PCRF and it receives an ARR command , the PCRF shall accept the request and only update the context(s) that are not in the process of being removed.

3.3 UPCON USE SCENARIOS

In this section four scenarios are described to show how UPCON operates to send user-plane congestion information to the PCRF in order to mitigate congestion.

3.3.1 SCENARIO 1 – A NEW UE ENTERS A CELL

In this scenario, presented in Figure 3.7, a new UE associates to a congested (E)UTRAN cell. The UE is accessing HD video. Once the RCAF receives this information it must send a RUCI notifying the PCRF of this event.

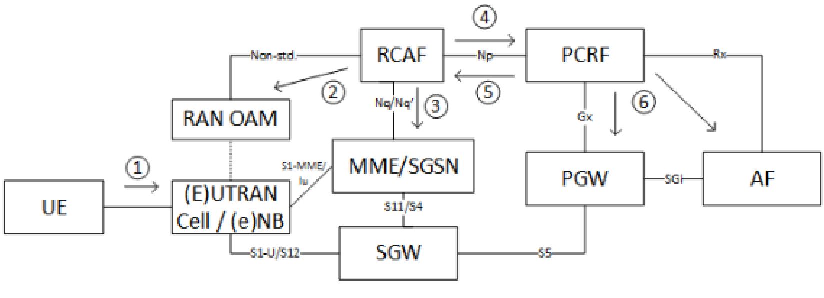


Figure 3.7: UPCON Scenario 1 - A new UE enters a congested cell

1. A UE connects to a cell changing the cell congestion level.

2. The RCAF retrieves congestion information from the RAN OAM serving that cell. For UTRAN the RCAF receives the list of affected IMSIs.

3. The RCAF selects the MME or SGSN serving that congested cell and retrieves from the MME or SGSN the IMSI and APN information of the users served by that congested cell.
4. The RCAF detects that a new user entered the congested area and prepares a NRR command and sends the Non-aggregated RUCI report to the PCRF. The RCAF stores UE context for that UE.
5. The PCRF stores the congestion information in the NRR command and replies with a NRA command.
6. The PCRF takes the congestion information into account for policy decisions. Policies are sent to the PGW and AF. According to the user profile, the bandwidth shall be adjusted. If the user is premium, users of lower profiles within that cell will have their QoS reduced.

3.3.2 Scenario 2 – A UE leaves a congested cell

In this scenario, presented in Figure 3.8, a UE leaves a congested cell and the cell congestion level changes.

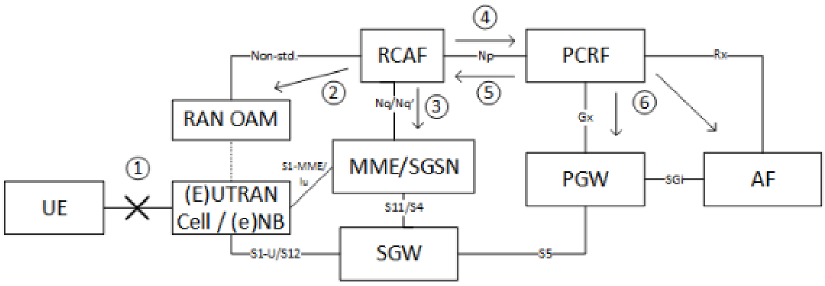


Figure 3.8: UPCON Scenario 2 - A UE leaves a congested cell

1. A UE leaves a congested cell changing the cell's congestion level.
2. The RCAF retrieves congestion information from the RAN OAM serving that cell. For UTRAN the RCAF receives the list of affected IMSIs.
3. The RCAF selects the MME or SGSN serving the congested cell and retrieves from the MME or SGSN the IMSI and APN information of the users served by that congested cell.
4. The RCAF detects that a user is no longer connected to a congested area and prepares a NRR command and sends the Non-aggregated RUCI report to the PCRF indicating that the UE is no longer congested. At the same time, the RCAF detects that cell congestion level changed so it prepares one or more ARR commands and sends those Aggregated RUCI reports to the PCRF. The RCAF updates the UE context for all these UEs.
5. The PCRF stores the congestion information in the NRR command and replies with a NRA command. The PCRF stores the congestion information in the ARR command and replies with an ARA command.
6. The PCRF takes the congestion information into account for policy decisions. Policies are sent to the PGW and AF. As congestion level changes QoS must be adapted (i.e., increase QoS levels of currently attached users to take advantage of resources that were released).

3.3.3 SCENARIO 3 – A UE MOVES BETWEEN CELLS

In this scenario, presented in Figure 3.9, a UE leaves the congested cell A and moves to cell B served by the same RCAF.

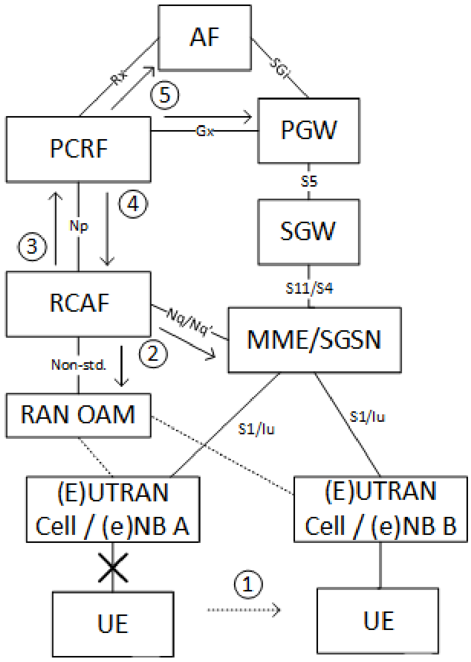


Figure 3.9: UPCON Scenario 3 - A UE moves between Cells

- Case A) Cell B is congested.
  1. A UE leaves a congested cell A and moves to the congested cell B.
  2. The RCAF serving the congested cells retrieves congestion information from the RAN OAM. The RCAF selects the MME(s) or SGSN(s) serving the congested area. The RCAF retrieves from the MME or SGSN the IMSI and APN information of the users served by the congested cells.
  3. The RCAF detects that a user moved from cell A to the congested cell B and, if the location change reporting is enabled, prepares a NRR command and sends the Non-aggregated RUCI report to the PCRF with location change indication. The RCAF updates UE context for that UE.
  4. The PCRF stores the congestion information in the NRR command and replies with a NRA command with optional reporting restrictions for that UE.
  5. The PCRF takes the congestion information into account for policy decisions.
- Case B) Cell B is not congested
  1. A UE leaves a congested cell A and moves to the congested cell B.
  2. The RCAF serving the congested cell retrieves congestion information from the RAN OAM. The RCAF selects the MME(s) or SGSN(s) serving the congested area. The RCAF retrieves from the MME or SGSN the IMSI and APN information of the users served by that congested cell.
  3. The RCAF detects that a user is no longer connected in a congested area it serves and prepares a NRR command and sends the Non-aggregated RUCI report to the PCRF indicating that the UE is no longer congested. The RCAF updates UE context for that UE. (The RCAF is only notified that the UE moved to cell B if it became congested.)

4. The PCRF stores the congestion information in the NRR command and replies with a NRA command.
5. The PCRF takes the congestion information into account for policy decisions.

3.3.4 Scenario 4 – A UE moves between RCAFs

In this scenario, presented in Figure 4, a UE moves from cell A served by RCAF 1 to cell B served by RCAF 2.

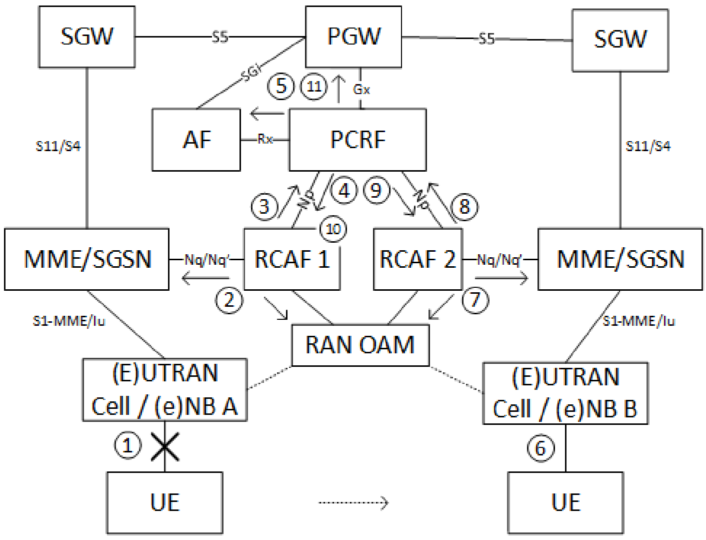


Figure 3.10: UPCON Scenario 4 - A UE moves between RCAFs

- Case A) Cell B is congested.
  1. A UE leaves Congested cell A served by RCAF 1.
  2. RCAF 1 retrieves congestion information from the RAN OAM serving the congested cell. The RCAF selects the MME(s) or SGSN(s) serving the congested area. The RCAF retrieves from the MME or SGSN the IMSI and APN information of the users served by the congested cell.
  3. The RCAF 1 detects that a user is no longer connected in the congested area it serves and prepares a NRR command and sends the Non-aggregated RUCI report to the PCRF indicating that the UE is no longer congested. The RCAF updates the UE context for that UE.
  4. The PCRF stores the congestion information in the NRR command and replies with a NRA command.
  5. The PCRF takes the congestion information into account for policy decisions.
  6. The UE enters the congested cell B who is covered by RCAF 2.
  7. RCAF 2 retrieves congestion information from the RAN OAM serving the congested cell. The RCAF selects the MME(s) or SGSN(s) serving the congested area. The RCAF retrieves from the MME or SGSN the IMSI and APN information of the users served by the congested cell.
  8. RCAF 2 detects that a new user entered the congested area it serves and prepares a NRR command and sends the Non-aggregated RUCI report to the PCRF. The RCAF stores UE context for that UE.

9. The PCRF receives RUCI for the same UE so it updates the congestion information received in the NRR command and replies with a NRA command.
10. The UE context in RCAF 1 shall be released so the PCRF sends via Np to RCAF 1 a MUR command with a “Release Context” indication. The RCAF releases the UE context and replies with a MUR command.
11. The PCRF takes the new congestion information into account for policy decisions.

- Case B) Cell B is not congested
  1. A UE leaves Congested cell A served by RCAF 1.
  2. RCAF 1 retrieves congestion information from the RAN OAM serving the congested cell. The RCAF selects the MME(s) or SGSN(s) serving the congested area. The RCAF retrieves from the MME or SGSN the IMSI and APN information of the users served by the congested cell.
  3. The RCAF 1 detects that the user is no longer connected in the congested area it serves and prepares a NRR command and sends the Non-aggregated RUCI report to the PCRF indicating that the UE is no longer congested. The RCAF updates UE context for that UE.
  4. The PCRF stores the congestion information in the NRR command and replies with a NRA command.
  5. The PCRF takes the congestion information into account for policy decisions.
  6. The UE enters the not congested cell B who is covered by RCAF 2.
  7. The RCAF serving the congested cell retrieves congestion information from the RAN OAM. As cell B is not congested the RCAF 2 does not request UE information. The PCRF is notified that the UE entered cell B via Gx.

3.4 NEW PROTOCOLS

3.4.1 RCAF-PCRF: NP PROTOCOL

To be able to implement congestion mitigation policies based on the RAN user plane congestion level, the PCRF must receive the RUCI from the RCAF.

The RUCI includes:

- The IMSI identifying the UE impacted by congestion;
- The eNB identifier, ECGI or SAI identifying the eNB, E-UTRAN cell or Service Area, respectively, serving the EU;
- The APN for which congestion information is reported; and
- The Congestion level or an indication of the "no congestion" state.

The operator shall be able to implement (optionally) RUCI reporting restrictions on a per-UE and per-APN basis to enable or disable reporting. The PCRF adds, updates and removes reporting restrictions on the RCAF through the Np. Restrictions may be used to limit the reporting, defining sets of non-overlapping congestion levels, such that the report indicates the congestion is within a given set instead of the level itself.

Two types of messages are used on Np to send congestion information from RCAF to PCRF:

- Non-aggregated RUCI report messages, which are sent on a per-UE and per-APN basis using DRA routing. The IMSI and the APN can be used to route messages.
- Aggregated RUCI report messages, which are sent between an RCAF and PCRF and contain congestion information for multiple UEs. A logical PCRF Id is allocated for these aggregate messages, which determine the destination of the message.

The decision regarding which information should be contained in a single aggregated RUCI report message out of the UEs with a given logical PCRF Id is out of the scope of 3GPP specifications, e.g. the RCAF may aggregate information only for a given cell or eNB into a single message. Alternatively, the RCAF may wait for a configurable period of time to aggregate information from multiple cells or eNBs into a single message.

The amount of RUCI updates may be limited by configuring a minimum time between RUCI updates in the RCAF (e.g. when only the identifier of the congested cell serving the UE has changed). This configuration is expected to take both the required accuracy as well as the acceptable signaling amount into account.

A RCAF shall serve a certain geographical area and cannot by itself detect if a UE moved to an area handled by another RCAF. When a RCAF indicates that a given UE is not affected by congestion, this does not exclude that another RCAF may report that the same UE experiences congestion. To ensure consistent operation in case of UE mobility, the PCRF stores information regarding the new RCAF and sends a message to the old RCAF to explicitly release context corresponding to the given UE and given APN, including any reporting restrictions. The criteria used for congestion detection is outside the scope of the 3GPP specifications.

The Np protocol used in the exchange between the PCRF and the RCAF is defined as a vendor specific Diameter application, where the vendor is 3GPP (Vendor-Id shall be set to 10415) with an Application-ID not yet defined. Regarding the Diameter protocol, the RCAF acts as the client (reporting the RUCI) and the PCRF acts as the server (handling the RUCI)

3.4.1.1 NP MESSAGES

The Np messages (Diameter Commands) exchanged between the RCAF and the PCRF with their respective AVPs are detailed below. They are defined according to the ABNF specification [51] as stated in clause 3.2 of [52].

1. NRR command. It is sent by the RCAF to the PCRF as part of the Non-aggregated RUCI report procedure and has the following message format.

```
<NR-Request> ::= <Diameter Header: 8388720, REQ, PXY >
          < Session-Id >
          [ DRMP ]
          { Vendor-Specific-Application-Id }
          { Auth-Session-State }
          { Origin-Host }
          { Origin-Realm }
          { Destination-Realm }
          [ Destination-Host ]
          [ Origin-State-Id ]
```

```
[ Subscription-Id ]
[ Called-Station-Id ]
[ Congestion-Level-Value ]
[ Congestion-Level-Set-Id ]
[ Congestion-Location-Id ]
[ OC-Supported-Features ]
[ RCAF-Id ]
*[ Proxy-Info ]
*[ Route-Record ]
*[ Supported-Features ]
*[ AVP ]
```

2. NRA command. It is sent by the PCRF to the RCAF as part of the Non-aggregated RUCI report procedure and has the following message format.

```
<NR-Answer> ::=  < Diameter Header: 8388720, PXY >
          < Session-Id >
          [ DRMP ]
          { Vendor-Specific-Application-Id }
          { Auth-Session-State }
          { Origin-Host }
          { Origin-Realm }
          [ Result-Code ]
          [ Experimental-Result ]
          [ Error-Message ]
          [ Error-Reporting-Host ]
          *[ Failed-AVP ]
          [ OC-Supported-Features ]
          [ OC-OLR ]
          [ Reporting-Restriction ]
          [ Conditional-Restriction ]
          [ RUCI-Action ]
          *[ Congestion-Level-Definition ]
          [ PCRF-Address ]
          [ Origin-State-Id ]
          *[ Redirect-Host ]
          [ Redirect-Host-Usage ]
          [ Redirect-Max-Cache-Time ]
          *[ Proxy-Info ]
          *[ Supported-Features ]
          *[ AVP ]
```

3. ARR command. It is sent by the RCAF to the PCRF as part of the Aggregated RUCI report procedure and has the following message format.

```
<AR-Request> ::= <Diameter Header: 8388721, REQ, PXY >
        < Session-Id >
        [ DRMP ]
        { Vendor-Specific-Application-Id }
```

```
{ Auth-Session-State }
{ Origin-Host }
{ Origin-Realm }
{ Destination-Realm }
[ Destination-Host ]
[ Origin-State-Id ]
*[ Aggregated-RUCI-Report ]
[ OC-Supported-Features ]
*[ Proxy-Info ]
*[ Route-Record ]
*[ Supported-Features ]
*[ AVP ]
```

4. ARA command. It is sent by the PCRF to the RCAF as part of the Aggregated RUCI report procedure and has the following message format.

```
<AR-Answer> ::= < Diameter Header: 8388721, PXY >
        < Session-Id >
        [ DRMP ]
        { Vendor-Specific-Application-Id }
        { Auth-Session-State }
        { Origin-Host }
        { Origin-Realm }
        [ Result-Code ]
        [ Experimental-Result ]
        [ Error-Message ]
        [ Error-Reporting-Host ]
        *[ Failed-AVP ]
        [ Origin-State-Id ]
        [ OC-Supported-Features ]
        [ OC-OLR ]
        *[ Redirect-Host ]
        [ Redirect-Host-Usage ]
        [ Redirect-Max-Cache-Time ]
        *[ Proxy-Info ]
        *[ Supported-Features ]
        *[ AVP ]
```

5. MUR command. It is sent by the PCRF to the RCAF in order to request congestion reporting restrictions for a specific UE and PDN ID and has the following message format.

```
< Modify-Uecontext-Request > ::= < Diameter Header: 8388722, REQ, PXY >
              < Session-Id >
              [ DRMP ]
              { Vendor-Specific-Application-Id }
              { Auth-Session-State }
              { Origin-Host }
              { Origin-Realm }
              { Destination-Realm }
```

```
{ Destination-Host }
[ Origin-State-Id ]
[ Subscription-Id ]
[ Called-Station-Id ]
[ OC-Supported-Features ]
[ Reporting-Restriction ]
[ Conditional-Restriction ]
[ RUCI-Action ]
*[ Congestion-Level-Definition ]
*[ Proxy-Info ]
*[ Route-Record ]
*[ AVP ]
```

6. MUA command. It is sent by the RCAF to the PCRF as a response to the request for congestion reporting restrictions to a specific UE and PDN ID and has the following message format.

```
< Modify-Uecontext-Answer > ::= < Diameter Header: 8388722, PXY >
        < Session-Id >
        [ DRMP ]
        { Vendor-Specific-Application-Id }
        { Auth-Session-State }
        { Origin-Host }
        { Origin-Realm }
        [ Result-Code ]
        [ Experimental-Result ]
        *[ Failed-AVP ]
        [ Origin-State-Id ]
        [ OC-Supported-Features ]
        [ OC-OLR ]
        *[ Redirect-Host ]
        [ Redirect-Host-Usage ]
        [ Redirect-Max-Cache-Time ]
        *[ Proxy-Info ]
        *[ AVP ]
```

3.4.1.2 NP SPECIFIC AVPS

New Diameter AVPs were defined to be used in Np and their Vendor-Id header is 3GPP (10415). Table 3.1 presents their AVP Code values, types, possible flag values, whether or not the AVP may be encrypted and what access types (e.g. 3GPP-EPS, etc.) the AVP is applicable to.

Table 3.1: Np specific Diameter AVPs

Attribute Name	AVP Code	Value Type	AVP Flag rules (1)			May Encr.	Applicability (3)
			Must	May	Must not		
Aggregated-Congestion-Info	4000	Grouped	V, M	P		Y	
Aggregated-RUCI-Report	4001	Grouped	V, M	P		Y	
Congestion-Level-Definition	4002	Grouped	V	P	M	Y	ReportRestriction
Congestion-Level-Range	4003	Unsigned32	V	P	M	Y	ReportRestriction
Congestion-Level-Set-Id	4004	Unsigned32	V	P	M	Y	ReportRestriction
Congestion-Level-Value	4005	Unsigned32	V, M	P		Y	
Congestion-Location-Id	4006	Grouped	V	P	M	Y	ReportRestriction
Conditional-Restriction	4007	Unsigned32	V	P	M	Y	ReportRestriction
eNodeB-Id	4008	OctecString	V, M	P		Y	
IMSI-List	4009	OctetString	V, M	P		Y	
RCAF-Id	4010	DiameterIdentity	V, M	P		Y	
Reporting-Restriction	4011	Unsigned32	V	P	M	Y	ReportRestriction
RUCI-Action	4012	Unsigned32	V	P	M	Y	ReportRestriction

NOTE 1: The AVP header bit denoted as ‘M’, indicates whether support of the AVP is required. The AVP header bit denoted as ‘V’, indicates whether the optional Vendor-ID field is present in the AVP header. ‘P’ bit means Protected and indicates the need for encryption for end-to-end security. Further details are presented in IETF RFC 3588[52].

NOTE 2: The value types are defined in IETF RFC 3588[52].

NOTE 3: The “ReportRestriction” feature is an optional feature that indicates in the Supported-Features AVP the support of reporting restriction by the RCAF and the PCRF and shall be included in every NRR and NRA commands.

The description of the Np AVPs and their format are now presented.

1. **Aggregated-Congestion-Info AVP** (type Grouped) - contains a list of user IDs identified by IMSI (in the IMSI-List AVP) and optionally (if the location reporting is enabled) the congestion location ID in which the list of user IDs are located (in the Congestion-Location-Id AVP).

```
Aggregated-Congestion-Info ::= < AVP Header: 4000 >
              [ Congestion-Location-Id ]
              [ IMSI-List ]
              *[ AVP ]
```

2. **Aggregated-RUCI-Report AVP** (type Grouped) – contains for a set of users which have the same PCRF for the same PDN ID the congestion level value (in the Congestion-Level-Value AVP) when no reporting restrictions were provided for the user ID and PDN ID or otherwise the congestion level set identifier (in the Congestion-Level-Set-Id AVP) if the PCRF provided the reporting restriction earlier for the user ID and PDN ID. The Called-Station-Id AVP contains the PDN ID and the user list and location are included in the Aggregated-Congestion-Info AVP.

```
Aggregated-RUCI-Report ::= < AVP Header: 4001 >
           1*{ Aggregated-Congestion-Info }
           [ Called-Station-Id ]
           [ Congestion-Level-Value ]
```

```
        [ Congestion-Level-Set-Id ]
        *[ AVP ]
```

3. **Congestion-Level-Definition AVP** (type Grouped) – is used to define a congestion level set and corresponding congestion level(s) to be used by the RCAF for a specific UE and PDN ID at congestion reporting. Reporting restriction apply for the specific UE and PDN ID when this AVP is present in the MUR command. It must contain Congestion-Level-Set-Id and Congestion-Level-Range AVPs.

```
  Congestion-Level-Definition ::= < AVP Header: 4002 >
        { Congestion-Level-Set-Id }
        { Congestion-Level-Range }
        *[ AVP ]
```

4. **Congestion-Level-Range AVP** (type Unsigned32) – is used to indicate the list of congestion level(s) bound to a certain congestion level set, between the PCRF and the RCAF. This AVP contains a bit mask where each bit has the meaning defined in Table 3.2. Bit 0 is the least significant bit.

Table 3.2: Congestion-Level-Range AVP

Bit	Name	Description
0	No congestion	This bit, when set, indicates that the RCAF shall report the corresponding congestion level set ID to the PCRF when there is no congestion for a certain UE and PDN ID.
1	Congestion level 1	This bit, when set, indicates that the RCAF shall report the corresponding congestion level set ID to the PCRF when congestion level 1 is reached for a certain UE and PDN ID.
1+n	Congestion level 1+n	This bit, when set, indicates that the RCAF shall report the corresponding congestion level set id to the PCRF when congestion level 1+n is reached for a certain UE and PDN ID.

5. **Congestion-Level-Set-Id AVP** (type Unsigned32) – is used to indicate the congestion level set identifier between the PCRF and the RCAF.

6. **Congestion-Level-Value AVP** (type Unsigned32) – is used to indicate the congestion level associated with the user ID and PDN ID. The defined values are presented in Table 3.3, ranging from 0 to 31.

Table 3.3: Congestion-Level-Value AVP

Value	Meaning
0	No congestion - indicates that there is no congestion.
N	Congestion level n - the value n is an integer between 1 and 31 and indicates a congestion level. The value 1 is the lowest congestion level and value 31 is the highest congestion level.

7. **Congestion-Location-Id AVP** (type Grouped) − indicates the identifier of a congested location in which the UE is currently located per UE and APN. The congested location is either the eNodeB, E-UTRAN cell or the Service Area serving the UE. If an ECGI is included in the 3GPP-User-Location-Info AVP then the eNodeB ID within eNodeB-ID AVP shall not be included in this AVP.

```
Congestion−Location−Id ::= < AVP Header: 4006 >
        [ 3GPP−User−Location−Info ]
        [ eNodeB−Id ]
        ∗[ AVP ]
```

8. **Conditional-Restriction AVP** (type Grouped) - contains a bit mask, and indicates what conditional reporting restrictions apply i.e. if the UE location info is included in RUCI. Conditional reporting restrictions are applied when this AVP is provided. Bit 0 (the least significant bit), when set, indicates that the location information of the UE (in the 3GPP-User-Location-Info AVP) shall not be included in RUCI for reporting.

9. **eNodeB-ID AVP** (type OctetString) - indicates the eNodeB in which the UE is currently located. The AVP shall be coded as in subclause 8.51 of 3GPP TS 29.274 [53].

10. **IMSI-List AVP** (type OctetString) - contains a list of user IDs identified by IMSI. Each IMSI shall be encoded as TBCD digits as defined in ITU-T Recommendation E.212 [54]. An example of the encoding of the IMSI List as the data part of the OctetString AVP is given below in Figure 3.11. Octet 1 given below is the first octet of the data part of the IMSI-List AVP and it is assumed that N IMSIs are included in the IMSI-List AVP. The digit length of IMSI-1, IMSI-2, IMSI-3 and IMSI-n are 15, 14, 14, 15 respectively.

Octets	Bits 8 7 6 5	Bits 4 3 2 1
1	Number digit 2 of IMSI-1	Number digit 1 of IMSI-1
2	Number digit 4 of IMSI-1	Number digit 3 of IMSI1
…	…	…
8	1111	Number digit 15 of IMSI-1
9	Number digit 2 of IMSI-2	Number digit 1 of IMSI-2
10	Number digit 4 of IMSI-2	Number digit 3 of IMSI-2
…	…	…
15	Number digit 14 of IMSI-2	Number digit 13 of IMSI-2
16	1111	1111
17	Number digit 2 of IMSI-3	Number digit 1 of IMSI-3
18	Number digit 4 of IMSI-3	Number digit 3 of IMSI-3
…	…	…
23	Number digit 14 of IMSI-3	Number digit 13 of IMSI-3
24	1111	1111
…	…	…
(n-1)*8+1	Number digit 2 of IMSI-n	Number digit 1 of IMSI-n
(n-1)*8+2	Number digit 4 of IMSI-n	Number digit 3 of IMSI-n
…	…	…
(n-1)*8+8	1111	Number digit 15 of IMSI-n

Figure 3.11: IMSI-List-AVP encoding example

11. **RCAF-Id AVP** (type DiameterIdentity) - is used to contain the RCAF identity.

12. **Reporting-Restriction AVP** (type Unsigned32) - is used to indicate the applicable reporting restrictions between the PCRF and RCAF. Unconditional reporting is applied when this AVP is not provided. Table 3.4 presents the defined values for reporting restriction.

Table 3.4: Reporting-Restriction AVP Values

Value	Meaning	Usage
0	No reporting restriction	Used by the PCRF to indicate to the RCAF that there are no restrictions on congestion reporting for a specific user IDuser ID and PDN ID. This value shall not be used if congestion level definitions are included in the same command.
1	Conditional reporting restriction	Used by the PCRF to indicate to the RCAF that there are conditional restrictions on congestion reporting for a specific user IDuser ID and PDN ID.
2	Unconditional reporting restriction	Used by the PCRF to indicate to the RCAF that there are unconditional restrictions on congestion reporting for a specific user IDuser ID and PDN ID.

13. **RUCI-Action AVP** (type Unsigned32) - is used to disable or enable RUCI reporting between the PCRF and RCAF or to release the context for a specific user IDuser ID and PDN ID. Table 3.5 presents the defined values for the AVP.

Table 3.5: RUCI-Action AVP Values

Value	Meaning	Usage
0	Disable RUCI reporting	Used by the PCRF to indicate to the RCAF that RUCI reporting shall not be performed for a specific UE and PDN ID.
1	Enable RUCI reporting	Used by the PCRF to indicate to the RCAF that RUCI reporting shall be performed for a specific UE and PDN ID.
2	Release Context	Used by the PCRF to indicate to the RCAF that the context shall be released for a specific user ID and PDN ID.

3.4.2 RCAF TO MME/SGSN: NQ-AP PROTOCOL

The Nq interface is located between the RCAF and the MME as specified in 3GPP TS 23.401 – Clause 5.1.1.13 [35]. The Nq' interface is located between the RCAF and the SGSN as specified in 3GPP TS 23.060 – Clause 5.6.3.12 [55]. They both use the Nq-AP protocol stack shown in Figure 3.12, which uses a Type, Length and Value (encoding) (TLV) encoding of messages over an SCTP transport (IETF RFC 4960 [56]) and its registered port number is 36424. This protocol is specified in 3GPP TS 29.405 [48] and a description of the defined procedures, messages and Information Elements is presented below.

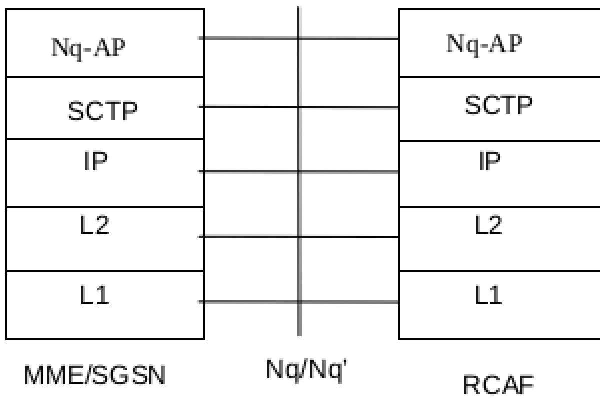


Figure 3.12: The Nq-AP protocol stack

3.4.2.1 IMSI AND APN INFORMATION RETRIEVAL PROCEDURE

It is used between the RCAF and the MME over the Nq interface. The RCAF uses this procedure to retrieve from the MME:

- The IMSIs of the UEs that are currently in ECM-CONNECTED state and having active E-RABs under a given eNodeB or E-UTRAN cell;
- For each IMSI the list of APNs currently having active PDN connections.

The RCAF selects the MME(s) serving a user plane congested eNodeB or an E-UTRAN cell based on the TAI of the cell or eNodeB that is congested in the user plane through the DNS procedure detailed in sub-clause 5.4 of 3GPP TS 29.303[50].

The RCAF sends a **NqAP-IMSI-APN-INFORMATION-REQUEST** message towards the MME(s) with:

- The Global eNodeB ID IE(s) if it requests the IMSI(s) and APN(s) for a given user plane congested area at an eNodeB level;
- The ECGI IE(s) if it requests the IMSI(s) and APN(s) for a given user plane congested area at an E-UTRAN cell level.

The MME reports in the **NqAP-IMSI-APN-INFORMATION-RESPONSE** message, for each eNodeB or E-UTRAN cell requested by the RCAF, that has at least one subscriber:

- the Macro eNodeB ID or Home eNodeB ID or the E-UTRAN cell ID in the RAN Entity value field of the RAN Entity Identifier IE for which the MME(s) report the IMSI and the APN information.
- all the subscribers that are currently in ECM-CONNECTED state and having active E-RABs, except the subscribers involved in an Emergency call, under the given eNodeB or E-UTRAN cell in the Subscriber-Information IE. Multiple instances of the Subscriber-Information IE shall be included if there are multiple subscribers under the given eNodeB or E-UTRAN cell.
- for each subscriber, the IMSI in the IMSI IE and APNs currently having active PDN connections in the APN IE. Multiple instances of the APN IE shall be included under a Subscriber-Information IE if the subscriber has active PDN connections towards multiple APNs.

No report is sent if the MME does not find any requested subscriber.

3.4.2.2 APN INFORMATION RETRIEVAL PROCEDURE

This procedure is used between the RCAF and the SGSN over the Nq' interface. The RCAF uses this procedure to retrieve from the SGSN the APNs that are currently having PDP contexts activated for each IMSI provided by the RCAF.

The RCAF selects the SGSN(s) serving a user plane congested SAI through the DNS procedure detailed in sub-clause 5.5 of 3GPP TS 29.303[50].

The RCAF sends an **NqAP-APN-INFORMATION-REQUEST** message towards the SGSN(s) with the list of IMSI(s) in the IMSI IE. Multiple instances of the IMSI IE shall be included if there are multiple subscribers for which APN information needs to be queried from the SGSN.

The SGSN reports in the **NqAP-APN-INFORMATION-RESPONSE** message, only if there is at least one subscriber's information to report the list of subscribers for which the APN information is provided by the SGSN, in the Subscriber-Information IE. Multiple instances of the Subscriber-Information IE shall be included if there are multiple subscribers for which the SGSN is reporting the APN information. Each instance of the Subscriber-Information IE contains:

- The IMSI of the subscriber encoded in the IMSI IE;

- The APNs currently having active PDP contexts with active RABs, encoded in the APN IE. Multiple instances of the APN IE shall be included under a Subscriber-Information IE if the subscriber has active PDP contexts towards multiple APNs. Emergency call APNs shall be excluded.
- If the subscriber does not have any PDP context with active RABs, then the SGSN shall not include that subscriber's information.

No report is sent if the SGSN does not find any requested subscriber.

3.4.2.3 PROTOCOL MESSAGES

The following messages are used in the Nq-AP protocol and the Information elements are detailed below.

NOTE: Information elements in these messages are either Mandatory (M) or Conditional (C). A Mandatory IE shall be included by the sending side and that the receiver diagnoses a "missing mandatory information element" error when the IE is not present. A Conditional IE shall be included by the sending entity if the conditions specified in the relevant protocol specification are met. The receiver shall check the conditions specified in the message to infer if the IE shall be expected and if so, when it is missing it shall abort the procedure.

1. **NqAP-IMSI-APN-INFORMATION-REQUEST Message**

   With Message type value 1, it is sent by the RCAF to the MME to request the list of IMSIs and APNs activated for each IMSI under a given RAN Identifier (E-UTRAN cell Id or the eNodeB Id) and the message content is shown in Table 3.6. Multiple instances of the RAN Entity Identifier IE shall be included if the RCAF is requesting the IMSI and APN information from MME for multiple eNodeB(s) or E-UTRAN Cell Ids.

   Table 3.6: NqAP-IMSI-APN-INFORMATION-REQUEST message content

Information element	IE Type	Presence	Format	Length
Message type	Message type	M	V	1
RAN Entity Identifier	RAN Entity Identifier	M	TLV	11-13

2. **NqAP-IMSI-APN-INFORMATION-RESPONSE Message**

   With Message type value 2, it is sent by the MME to the RCAF in response to the NqAP-IMSI-APN-INFORMATION-REQUEST message and its content is presented in Table 3.7. The RAN-Associated-Information IE shall be encoded if the cause value in Cause IE is "Request Accepted" and multiple instances of it shall be included if the RCAF had requested IMSI and APN information for multiple eNodeB Id(s) or EUTRAN Cell Id(s) and if the particular instance of this message could find subscribers under some or all of these requested eNodeB Id(s) or EUTRAN Cell Id(s). The MME Name IE shall be included if the cause value in Cause IE is anything other than "Request Accepted".

Table 3.7: NqAP-IMSI-APN-INFORMATION-RESPONSE message content.

Information element	IE Type	Presence	Format	Length
Message type	Message type	M	V	1
Cause	Cause	M	TLV	4
RAN Associated Information	RAN-Associated-Information	C	TLV	Variable
MME Name	Fully Qualified Domain Name (FQDN)	C	TLV	Variable

3. **NqAP-APN-INFORMATION-REQUEST Message**

   With Message type value 3, it is sent by the RCAF to the SGSN to request the list of APNs activated for a given set of IMSI(s) and Table 3.8 shows its content. Multiple instances of the IMSI IE shall be included if the RCAF is requesting the APN information from SGSN for multiple IMSIs.

Table 3.8: NqAP-APN-INFORMATION-REQUEST message content

Information element	IE Type	Presence	Format	Length
Message type	Message type	M	V	1
IMSI	IMSI	M	TLV	6-11

4. **NqAP-APN-INFORMATION-RESPONSE Message**

   With Message type value 4, it is sent by the SGSN to the RCAF in response to the NqAP-APN-INFORMATION-REQUEST message and Table 3.9 shows its content. The Subscriber-Information IE shall be encoded if the cause value in Cause IE is "Request Accepted". Multiple instances of the Subscriber-Information IE shall be included if the RCAF requests APN information for multiple IMSI(s) and if the particular instance of the NqAP-APN-INFORMATION-RESPONSE message carries more than one subscriber's information. The SGSN Name IE shall be included if the cause value in Cause IE is anything other than "Request Accepted".

Table 3.9: NqAP-APN-INFORMATION-RESPONSE message content

Information element	IE Type	Presence	Format	Length
Message type	Message type	M	V	1
Cause	Cause	M	TLV	4
Subscriber Information	Subscriber-Information	C	TLV	Variable
SGSN Name	Fully Qualified Domain Name (FQDN)	C	TLV	Variable

3.4.2.4 NQ-AP INFORMATION ELEMENTS

These are the information elements used by the Nq-AP protocol.

The Message Type uniquely identifies the message being sent. It is a single octet information element, mandatory in all messages and its value part is presented in Table 3.10.

Table 3.10: Message type information element

Message Type value	Message type
0	Reserved
1	NqAP-IMSI-APN-INFORMATION-REQUEST
2	NqAP-IMSI-APN-INFORMATION-RESPONSE
3	NqAP-APN-INFORMATION-REQUEST
4	NqAP-APN-INFORMATION-RESPONSE
5-255	Spare. For future use.

A Nq-AP message may contain several information elements. With the exception of the Message type IE all the other Information Elements are TLV coded. The Nq-AP information element type values are specified in the Table 3.11 and their generic format is depicted in Figure 3.13. Their coding is presented below according to 3GPP TS 29.405[48].

Table 3.11: Information Element types for Nq-AP

IE Type value (Decimal)	Information elements
0	Reserved
1	International Mobile Subscriber Identity (IMSI)
2	Access Point Name (APN)
3	RAN Entity Identifier
4	Subscriber-Information
5	RAN Associated Information
6	Cause
7	Fully Qualified Domain Name (FQDN)
8-255	Spare. For future use.

Octets	Bits							
	8	7	6	5	4	3	2	1
1	Type = xxx (decimal)							
2 to 3	Length = n							
4 to (n+3)	IE specific data or content of a grouped IE							

Figure 3.13: Information Element Format

1. **International Mobile Subscriber Identity (IMSI)**

   This IE represents the IMSI of the subscriber. The sending entity copies the value part of the IMSI into octets 4 to (n+3) of the IMSI IE encoded as TBCD digits as shown in Figure 3.14. The maximum number of digits is 15.

Octets	Bits							
	8	7	6	5	4	3	2	1
1	IE Type = 1(decimal)							
2 to 3	Length = n							
4	Number digit 2				Number digit 1			
5	Number digit 4				Number digit 3			
...	...				...			
n+3	Number digit m				Number digit m-1			

Figure 3.14: IMSI IE encoding

2. **Access Point Name (APN)**

   This IE (in Figure 3.15) specifies the APN towards which a subscriber has active PDN connection.

Octets	Bits							
	8	7	6	5	4	3	2	1
1	IE Type = 2(decimal)							
2 to 3	Length = n							
4 to (n+3)	Access Point Name (APN)							

Figure 3.15: APN IE encoding

3. **RAN Entity Identifier**

   This IE represents the location granularity for which subscriber information is provided. It shall either be a Macro eNodeB ID or a Home eNodeB ID or an E-UTRAN Cell Global Identity. Its format is shown in Figure 3.16.

<table>
<tr><th></th><th colspan="8">Bits</th></tr>
<tr><th>Octets</th><th>8</th><th>7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th></tr>
<tr><td>1</td><td colspan="8">IE Type = 3 (decimal)</td></tr>
<tr><td>2 to 3</td><td colspan="8">Length = n</td></tr>
<tr><td>4</td><td colspan="8">RAN Entity Type</td></tr>
<tr><td>5 to (n+4)</td><td colspan="8">RAN Entity Value</td></tr>
</table>

Figure 3.16: RAN Entity Identifier IE

The RAN Entity Type values are given below in Table 3.12.

Table 3.12: RAN Entity Type values and their meanings

RAN Entity Type	RAN Entity Value (Decimal)
Macro eNodeB ID	0
Home eNodeB ID	1
E-UTRAN Cell Global Identity	2
Spare	3 to 255

If the RAN Entity Type is a Macro eNodeB ID then the RAN Entity value shall represent a Macro eNodeB ID which is encoded as shown in Figure 3.17.

<table>
<tr><th></th><th colspan="8">Bits</th></tr>
<tr><th>Octets</th><th>8</th><th>7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th></tr>
<tr><td>5</td><td colspan="4">MCC digit 2</td><td colspan="4">MCC digit 1</td></tr>
<tr><td>6</td><td colspan="4">MNC digit 3</td><td colspan="4">MCC digit 3</td></tr>
<tr><td>7</td><td colspan="4">MNC digit 2</td><td colspan="4">MNC digit 1</td></tr>
<tr><td>8</td><td colspan="4">Spare</td><td colspan="4">Macro eNodeB ID</td></tr>
<tr><td>9 to 10</td><td colspan="8">Macro eNodeB ID</td></tr>
<tr><td>11 to 12</td><td colspan="8">Tracking Area Code (TAC)</td></tr>
</table>

Figure 3.17: Encoding of Macro eNodeB as the RAN Entity Value

The Macro eNodeB ID consists of 20 bits. Bit 4 of Octet 8 is the most significant bit and bit 1 of Octet 10 is the least significant bit. In the coding of the Macro eNodeB ID full hexadecimal representation shall be used.

If the RAN Entity Type is Home eNodeB ID then the Home eNodeB ID shall be encoded as the RAN Entity value as shown below in Figure 3.18.

	Bits							
Octets	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
5	MCC digit 2				MCC digit 1			
6	MNC digit 3				MCC digit 3			
7	MNC digit 2				MNC digit 1			
8	Spare				Home eNodeB ID			
9 to 11	Home eNodeB ID							
12 to 13	Tracking Area Code (TAC)							

Figure 3.18: Encoding of Home eNodeB as the RAN Entity Value

The Home eNodeB ID consists of 28 bits. Bit 4 of Octet 9 is the most significant bit and bit 1 of Octet 12 is the least significant bit. In the coding of the Home eNodeB ID full hexadecimal representation shall be used.

If the RAN Entity Type is E-UTRAN Cell Global Identity (ECGI) then the ECGI shall be encoded as the RAN Entity value as shown below in Figure 3.19.

	Bits							
Octets	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
5	MCC digit 2				MCC digit 1			
6	MNC digit 3				MCC digit 3			
7	MNC digit 2				MNC digit 1			
8	Spare				ECI			
9 to 11	ECI (E-UTRAN Cell Identifier)							

Figure 3.19: Encoding of ECGI as the RAN Entity Value

The E-UTRAN Cell Identifier (ECI) consists of 28 bits. The ECI field shall start with Bit 4 of octet 8, which is the most significant bit. Bit 1 of Octet 11 is the least significant bit. In the coding of the E-UTRAN cell identifier full hexadecimal representation shall be used.

4. **Subscriber-Information**

This IE is a grouped IE and it carries the information related to a subscriber (IMSI and APN) and is coded according to Figure 3.20.

Octet 1	IE Type = 4 (decimal)		
Octets 2 and 3	Length = n		
Information elements	**P**	**Condition / Comment**	**IE Type**
IMSI	M		IMSI
APN	M	If there are more than one APN for which PDN connections are activated for a subscriber, then each APN shall be encoded as a separate instance of the APN IE	APN

Figure 3.20: Subscriber Information IE encoding

5. **RAN Associated Information**

This IE is a grouped IE and it carries the information related to list of subscribers for a given RAN Entity Identifier and is coded according to Figure 3.21.

Octet 1	IE Type = 5 (decimal)		
Octets 2 and 3	Length = n		
Information elements	**P**	**Condition / Comment**	**IE Type**
RAN Entity Identifier	M		RAN Entity Identifier
Subscriber Information	M	If there are more than one subscriber under the given RAN Entity Identifier, then each subscriber's information shall be encoded as a separate instance of the Subscriber Information IE	Subscriber-Information

Figure 3.21: RAN Associated Information IE encoding

6. **Cause**

This IE shall be used to indicate the RCAF Success or reason for failure of processing a request message and shall be encoded as specified in Figure 3.22.

.	**Bits**							
Octets	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
1	IE Type = 6 (decimal)							
2 to 3	Length = n							
4	Cause value							
5	Type of the offending IE							

Figure 3.22: Cause IE encoding

Cause is a variable length IE, which may have either of the following two lengths values:

- If n = 1, then the Cause IE shall be 4 octets long. Therefore, octet 5 will not be present.
- If n = 2, then the Cause IE shall be 5 octets long.

The Cause value shall be included in a response message. In a response message, the Cause value indicates the acceptance or the rejection of the corresponding request message. The Cause value indicates the explicit reason for the rejection.

If the rejection is due to a mandatory IE or a verifiable conditional IE is faulty or missing, the offending IE shall be included in Octet 5.

Table 3.13 provides the various cause values.

Table 3.13: Cause values

Cause Type	Cause value	Meaning
	0	Reserved. Shall not be sent and if received the Cause shall be treated as an invalid IE.
Acceptance Cause	1	Request Accepted.
	2 to 15	Spare. This value range shall be used for acceptance cause values.
Protocol Errors	16	Mandatory IE incorrect.
	17	Mandatory IE missing.
	18	Conditional IE incorrect.
	19	Conditional IE missing.
	20	Invalid length.
	21 to 63	Spare. This value range shall be used by Cause values representing protocol errors.
Cause Codes Representing Status	64	MME/SGSN Congested.
	65 to 255	Spare. For Future Use.

7. **Fully Qualified Domain Name (FQDN)**

   This IE is coded as depicted in Figure 3.23. The FQDN field encoding shall be identical to the encoding of a FQDN within a DNS message of section 3.1 of IETF RFC 1035 [57] but excluding the trailing zero byte.

	Bits							
Octets	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
1	Type = 7 (decimal)							
2 to 3	Length = n							
4 to (n+3)	FQDN							

Figure 3.23: FQDN IE encoding

3.4.3 RCAF TO THE RAN OAM

The interface between the RCAF and the RAN OAM is not (yet) standardized. Furthermore, the criteria used in the detection of RAN user plane congestion (including the detection of congestion abatement) are outside the scope of 3GPP specifications (sub-clause 4.3.24.3 of 3GPP TS 23.401 [35]).

3.5 CHAPTER CONSIDERATIONS

This chapter addressed 3GPP's standardization effort in UPCON for Release 13. This work started with the definition of User-plane congestion and mitigation techniques and culminated with the introduction of a new network architecture element,the RCAF, and new interfaces and protocols. The Np interface stands between the RCAF and the PCRF and uses the Np protocol over Diameter while the Nq/Nq' stands between the RCAF and the MME or the Serving GPRS Support Node (SGSN) and uses the Nq-AP protocol over SCTP. These interfaces and protocols were detailed together with their signaling diagram. Some scenarios were included to represent the different ways UPCON can be used. It was noted that the standardized solution only involves Core Network elements in either the congestion reporting or management. This implementation leaves room to enhancement and interworking with other solutions to achieve a wider congestion management solution. Namely, an evolved UPCON where a RCAF enhanced to receive UE data could provide the PCRF with HotSpot 2.0 congestion information on a Wi-Fi APs, this information could then be used by the PCRF to send access policies to the UE via the ANDSF. In an environment where the MNO has deployed a Wi-Fi network the RCAF could receive congestion information from their AP.

CHAPTER 4

Congestion Management with ANDSF

This chapter details the network architecture and simulation environment envisioned in the context of the SMCon project selected as the starting point for the implementation work of this dissertation presented in Chapter 5. The purpose of said project was to access the contribution of the ANDSF mechanism in an operator network. To achieve this, Siafu[1] an open source context simulator was used to run a simulation as detailed. The chapter finishes presenting and analyzing the project results as well as the proposed simulation enhancements.

4.1 Introduction

One of the key mechanisms standardized by 3GPP for mobile traffic and connection management is the ANDSF. As seen in subsection 2.4.1, when deployed the ANDSF uses the S14 interface for a client/server interaction. The client (ANDSF-C) is installed in the UE and the server (ANDSF-S) is a Core Network element. Policies for access network discovery and selection can be exchanged either by request of the client in Pull Mode or in Push Mode when the server initiates the procedure. ANDSF gives a MNO the control on which networks the UE shall connect prioritizing them according to location, time of day and user profile. This mechanism is specially relevant in complementing 3GPP networks with Wi-Fi as an traffic offloading technology.

The simulation was developed to evaluate the deployment of ANDSF in an operator network covering a city with three kinds of users with different daily routines. Important metrics taken in consideration were signaling events (policy requests from the client), policy data exchanged and network access type.

The simulator includes a Graphical User Interface (GUI) window and allows the visualization of the simulation environment where agents move around a map according to the simulation parameters which

[1] http://siafusimulator.org/

was used for the simulation debugging. Siafu also permits disabling the GUI which was very important when what was wanted was the collection of context simulation data and not the visualization.

The simulation ANDSF-S policy generator selects policy based on information regarding user profile, location and movement speed but not on 3G/LTE cell congestion levels, therefore it was considered that extending the existing simulator with this feature would allow an evaluation of using congestion information provided by UPCON reporting to enhance the ANDSF policy generator.

4.2 SIMULATION SCENARIO

The simulation scenario developed in Siafu implemented the ANDSF network architecture (presented in subsection 2.4.1) and client/server communication as standardized in [35], [30] and [31].

The implementation follows these main sections:

- City Map: describes the maps used for agent movement, the Wi-Fi APs location and identifies the cells and LACs covering the city;
- Agent Profiles: describes the characteristics and use cases for each generated profile;
- Signaling Triggers: describes the defined triggers for ANDSF policy exchange;
- ANDSF server rule decision: describes the decision flow used to select the appropriate ANDSF policy;

The scenario, represented in Figure 4.1, explored the usage of Wi-Fi AP as an offloading technology with ANDSF as an UE access network manager. Decisions regarding the constants such as the number of places and agents generated used for the simulation were taken considering not only the reasonability of values in order to allow extrapolation but also on practicality as the simulation time extends from a few hours to several days as the numbers increase.

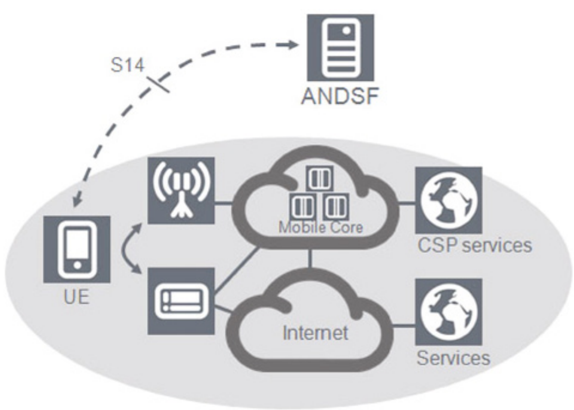
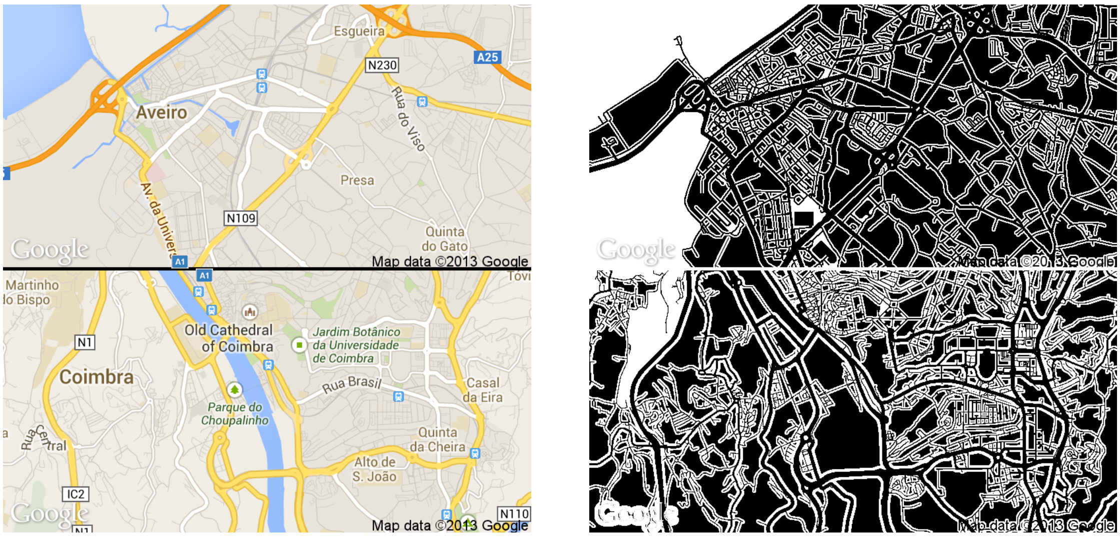


Figure 4.1: Simulation Scenario [58]

4.3 SIMULATION IMPLEMENTATION

4.3.1 CITY MAP

For this simulation maps for the cities of Aveiro and Coimbra obtained from Google Maps[2] (in Figure 4.2a) were used. These colored maps are only used as the background for the GUI simulation, it was necessary to provide the black and white map in Figure 4.2b indicating, in black, areas where the agents are able to move. Siafu is able to interpret the black regions as movable areas, for the user agents defined in the simulation.



(a) GUI background map

(b) Areas for agent movement

Figure 4.2: Simulation Maps for GUI background and Areas for agent movement

The 3GPP mobile cells covering the city maps are shown in 4.3 with the Mall cells in white. LAC-A2 corresponds to an area outside the city and LAC-C1 cover areas inside city areas.

[2]maps.google.pt

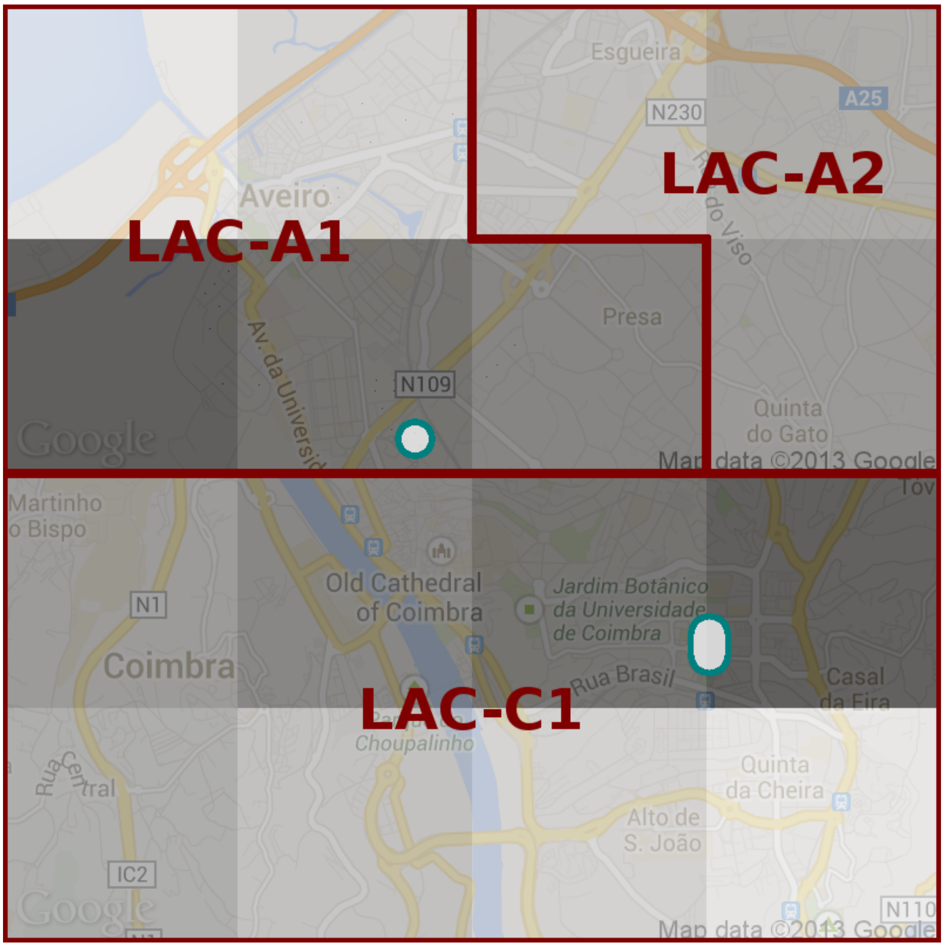
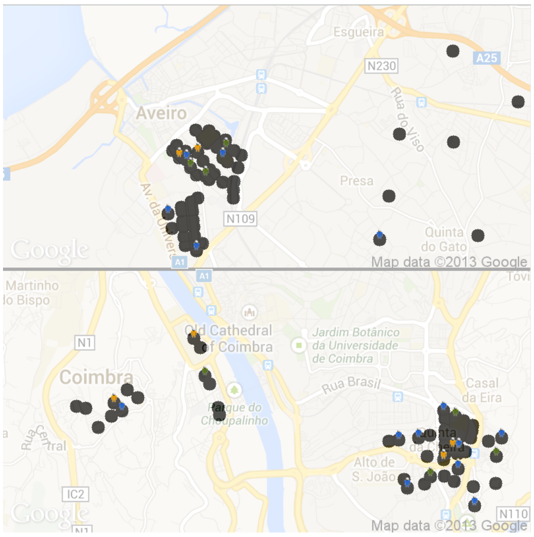


Figure 4.3: Cell and LAC Map coverage

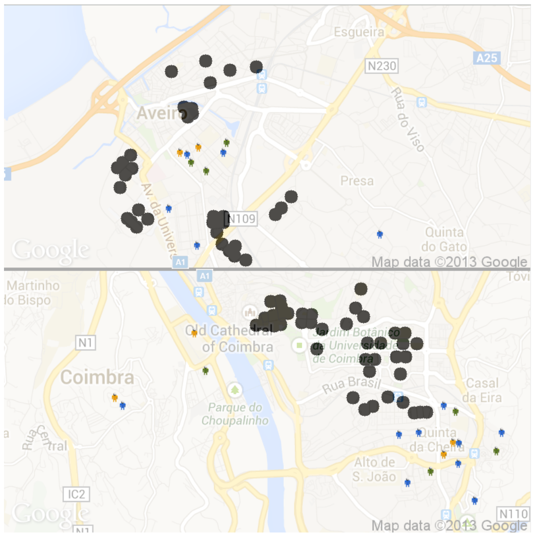
Four types of Wi-Fi APs were placed throughout each city as shown in Figure 4.4 where the black spots represent the Wi-Fi coverage points. The number of APs in each map follows the distribution presented in Table 4.1. It must be noted that the Mall and Premium APs represent multiple APs aggregated under a single SSID.

Table 4.1: Wi-Fi APs distribution

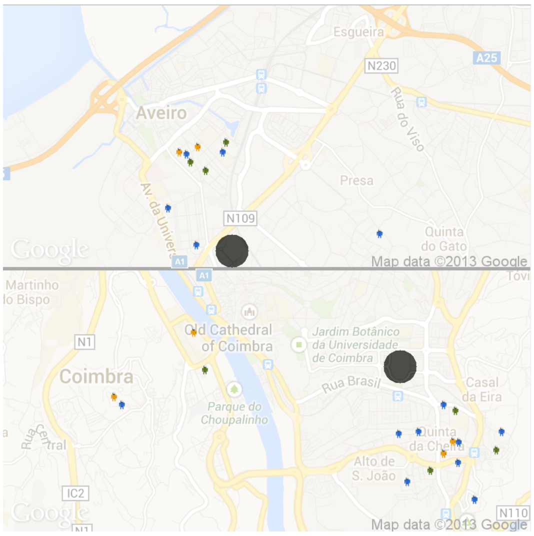
	Aveiro	Coimbra
Mall	1	1
Premium	1	1
Offices	28	37
Homes	51	50



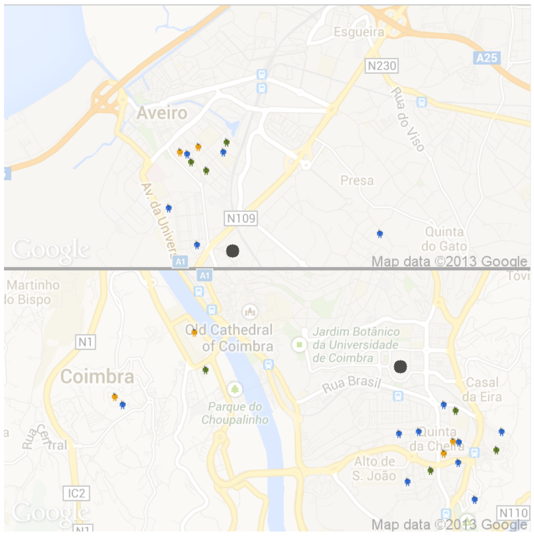
(a) Home APs



(b) Office APs



(c) Mall APs



(d) Premium APs

Figure 4.4: Wi-Fi Access Points of the simulation

4.3.2 AGENT PROFILES

The simulation implements 3 agent profiles (representing mobile users) each with various use cases which describe their whereabouts along the day. At the scenario generation, each agent receives randomly a specific Home Place and 35% of them receive a "Premium" profile.

The number of agents generated in each map is according to Table 4.2.

Table 4.2: Agents generated by profile

	Aveiro	Coimbra
Salary Man	100	200
Sales Man	50	100
Truck Driver	30	60

All agents are able to connect to Home, Office and Mall APs; Premium agents have a reserved AP at the Mall. Each agent profile has different characteristics:

Salary Man: Commutes from his designated Home Place to an Office Place and remains during work hours going to the Mall at lunch hours;

Sales Man: Works moving around its designated city and goes to the Mall at lunch hours;

Truck Driver: Usually on high mobility and is able to exchange city via the highway.

Table 4.3 presents all the defined use cases and Tables 4.4, 4.5 and 4.6 include the use cases related to each agent profile.

Table 4.3: Agent's use cases.

Use Case	Description
C1 - Home	User agent at home.
C2 - Comute	User agent commuting between two different places.
C3 - Work	User agent at work.
C4 - Lunch	User agent having lunch at the mall
C5 - Travelling	User agent travelling (drive around the city).
C6 - Stopping to eat	User agent stopping to eat (after driving).
C7 - Work outside	User agent working outside (walk around the city).

Table 4.4: Salary Man use cases.

Time	00:00 08:30	08:31 09:00	09:01 12:00	12:01 12:30	12:31 13:30	13:31 14:00	14:01 19:00	19:01 23:59
Use Case	C1 (home)	C2 (comute)	C3 (work)	C2 (comute)	C4 (lunch)	C2 (comute)	C3 (work)	Free time

Table 4.5: Sales Man use cases.

Time	00:00 08:30	08:31 09:00	09:01 12:30	12:31 13:30	13:31 18:00	18:01 23:59
Use Case	C1 (home)	C2 (comute)	C7 (work outside)	C4 (lunch)	C7 (work outside)	Free time

Table 4.6: Truck Driver use cases.

Time	00:00 06:00	06:01 12:00	12:01 14:00	14:01 17:00	17:01 23:59
Use Case	C1 (home)	C5 (travelling)	C4 (lunch)	C5 (travelling)	Free time

4.3.3 SIGNALING TRIGGERS

In the concept phase, a set of Signaling Triggers were defined to determine the need for an ANDSF-S response as the agents moved through the map. Triggers are either UE-based or Network-based if the response is requested by the UE or initiated by the network. Table 4.7 presents all the defined triggers, their type and description. At the development phase only the triggers in green were implemented.

Table 4.7: Simulation ANDSF-S Triggers

ID	Type	Trigger	Description
T1	UE[1]	ANDSF-C startup	Whenever the ANDSF-C starts-up, after initial connection.
T2	UE	On-demand User (button)	The User presses a button that triggers a request to S-ANDSF.
T3	N[2]	On-demand Server	On response, ANDSF-S sends information about when ANDSF-C should make a new request to update policies.
T4	N	Periodic	Weekly, Daily, every X hours, etc. The periodicity of ANDSF-C requests is in UE settings or included in every ANDSF-S responses.
T5	UE	Change of 3GPP RAC	Whenever the UE detects a RAC/MMEC change.
T6	UE	Change of 3GPP LAC	Whenever the UE detects a LAC/MMEGI change.
T7	UE	Change of 3GPP CI	Whenever the UE detects a CI change.
T8	UE	Change of 3GPP PLMN	Whenever the UE detects an operator change, i.e. PLMN-Id (on roaming).
T13	UE	Change of Mobility	Whenever the Network/UE detects a significant change in UE movement factor. ANDSF-S may send new policies to ANDSF-C or schedule for the next UE movement factor change.
T14	N	Network congestion	Whenever network congestion status changes, all UEs in the affected area, or some of them, receive new policies from ANDSF-S to adapt to the congestion status.
T15	UE	Wi-Fi signal strength	Whenever a UE is connected to a hotspot and its signal strength becomes weak.
T16	UE	Hotspots discovered	Whenever UE discovers more than a certain number of Hotspots, new policies from S-ANDSF can help hotspot selection.
T17	N	Network RAN fail	Whenever a network RAN fail is detected, all UE in affected area receive new policies from ANDSF-S with alternatives to the failed RAN.
T18	N	Network Hotspots update	Whenever the network updates the list of available Hotspots, ANDSF-S sends new policies to UE in affected areas.

[1] UE-based Trigger
[2] Network-based Trigger

4.3.4 SIGNALING

The signaling exchange, in Figure 4.5, between the ANDSF-Client in the agent's UE and the ANDSF-Server is performed only in Pull Mode as the only implemented triggers are UE-Based.
For each signaling trigger the Client sends via the S14 interface an Policy Request (Figure 4.5- 1) to the Server wich then selects the appropriate policy and sends (Figure 4.5- 2) the corresponding MO data to the UE.

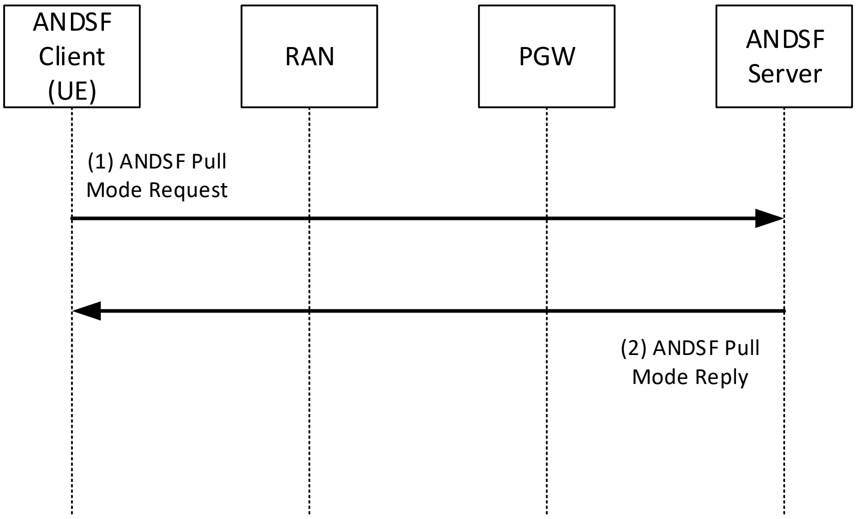


Figure 4.5: Implemented Signaling flow

ANDSF MO data in the XML tree structure format used in the S14 interface [31] were generated for each defined policy in order to find the amount of information exchanged. The rules engine was optimized to reply only with policy Ids in case of repeated policies.

4.3.5 ANDSF SERVER RULE DECISION

In this scenario the ANDSF-S rules engine follows the decision tree presented in Figure 4.6. The policy sent to the ANDSF-C is based on agent mobility, location and profile. A handover to Wi-Fi must always be performed whenever an user reaches a Wi-Fi AP except when in high mobility. This behavior may lead to poor user experience as the handovers may be unnecessary if the cell is not congested or if the Wi-Fi AP is congested.

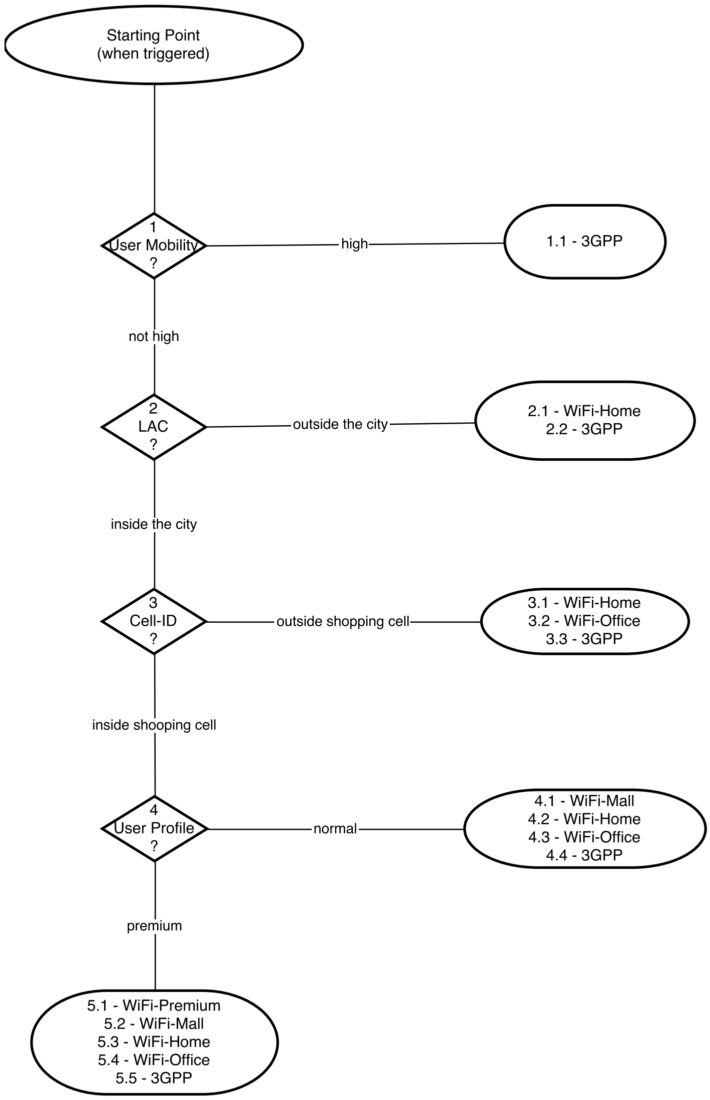


Figure 4.6: ANDSF-Server rule decision flow

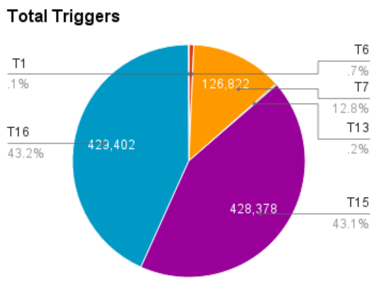
4.4 EVALUATION

The perform an evaluation of the developed scenario, the simulation was run three times, each representing a whole day. The simulator returned three log files formatted in Extensible Markup Language (XML) for each profile and use case with 1) the bytes exchanged for policy provision; 2) the number of triggers for policy request and 3) the time in each access time. The log files were then parsed with the python scripts in [59] creating Comma Separated Values (CSV) files allowing the data to be saved in a table format. Data was then treated and the results are presented in the following sections.

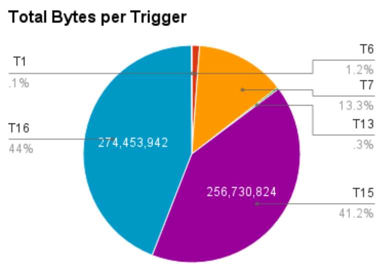
4.4.1 SIGNALING TRIGGERS

4.4.1.1 TRIGGER AMOUNT

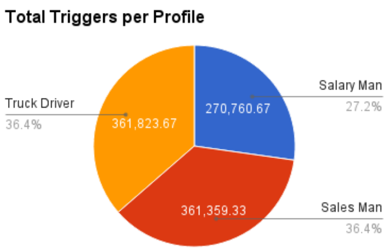
One of the main metrics analyzed, presented in Figure 4.7, is trigger information. The total number of triggers (in Figure 4.7a) and the total bytes of the policies sent by the trigger (in Figure 4.7b) appear equivalent in terms of magnitude and it is evident that the most significant part of policy provision is caused by *Trigger_15* (T15) and *Trigger_16* (T16) which correspond to loss and discovery of Wi-Fi access respectively. *Trigger_7* (T7) which corresponds to a change in Cell-ID is responsible for the next portion of data.



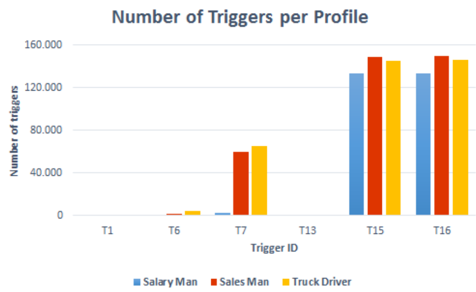
(a) Total Triggers



(b) Total Bytes



(c) Triggers per Profile



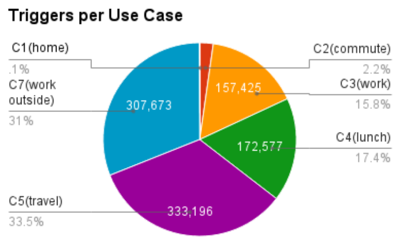
(d) Number of Triggers

Figure 4.7: Simulation Triggers

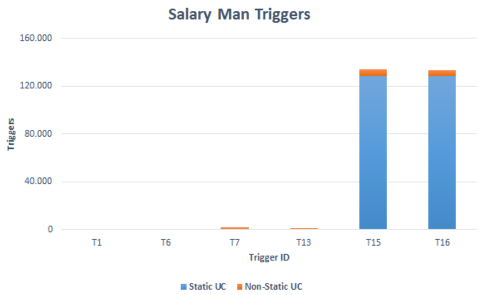
Regarding the distribution of triggers by each agent profile, Figure 4.7c reveals that the percentage of triggers of the *Sales Man* and the *Truck Driver* agents is equivalent, with both being slightly larger than *Salary Man*'s. This corresponds to the results in Figure 4.7d and the discrepancy is largely motivated by the numbers of *Trigger_7*.

4.4.1.2 IMPACT OF MOBILITY IN SIGNALING

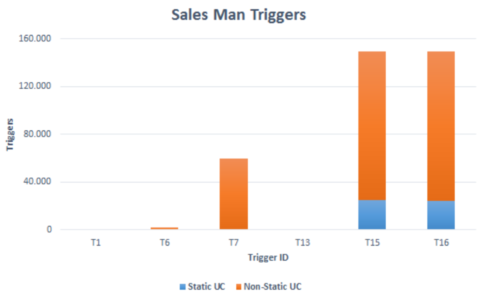
Another aspect taken in consideration was the Trigger distribution per Use Case and the impact of user mobility in triggers and is presented in Figure 4.8. Analyzing Figure 4.8a it is evident that *use_case_5* (travel) and *use_case_7* (work outside) are responsible for the most significant part of triggers followed by *use_case_4* (lunch) and *use_case_3* (work).



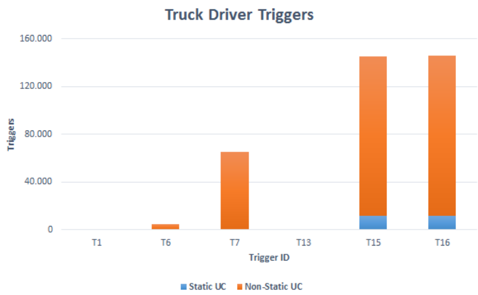
(a) Triggers per Use Case



(b) Salary Man Triggers



(c) Sales Man Triggers



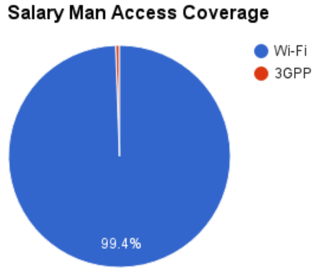
(d) Truck Driver Triggers

Figure 4.8: Impact of Mobility on Triggers

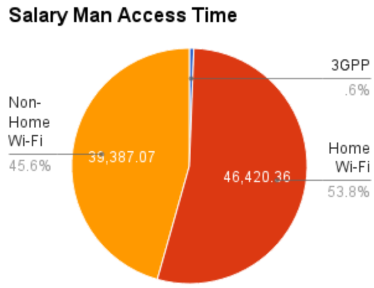
Agent mobility affects each agent profile in different ways, Figure 4.8b and Figure 4.8b indicate that the *Sales Man* and *Truck Driver* profiles have triggers mainly in non-static periods with prevalence for *Trigger_15* and *Trigger_16* but also significant numbers of *Trigger_7*. Figure 4.8b shows that the *Salary Man*'s triggers are almost entirely located during static periods.

4.4.2 Access Coverage and Access Time

Other important metrics analyzed were Access Coverage and Access Time which are presented in Figure 4.9 for the *Salary Man* and in Figure 4.10 for the *Sales Man* and the *Truck Driver* agents. Figure 4.9a and Figure 4.9b reveal that the *Salary Man* agent is covered 99.4% of its time by Wi-Fi divided almost evenly between Home and Non-Home Wi-Fi access.



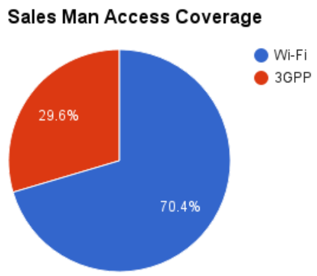
(a) Coverage Time



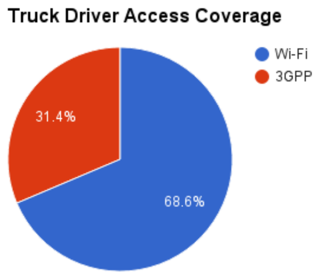
(b) Access Time

Figure 4.9: Salary Man Access Results

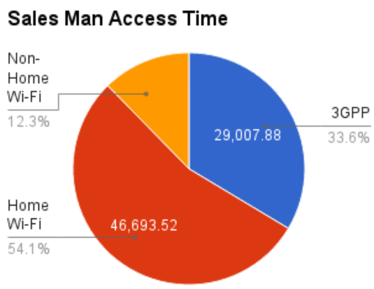
According to Figure 4.9a and Figure 4.10b, the *Sales Man* and the *Truck Driver* have similar access coverage times and regarding the access times (in Figure 4.9b and Figure 4.10d respectively), in both it surpasses 50% but in the *Sales Man* case there is more equal distribution between the 3GPP and Non-Home Wi-Fi. These results are in line with the agent's different characteristics.



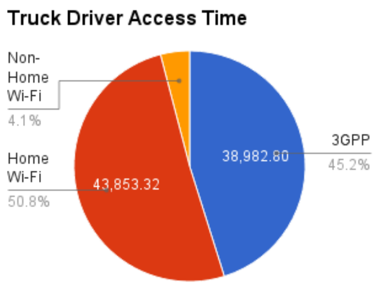
(a) Coverage Time



(b) Coverage Time



(c) Access Time

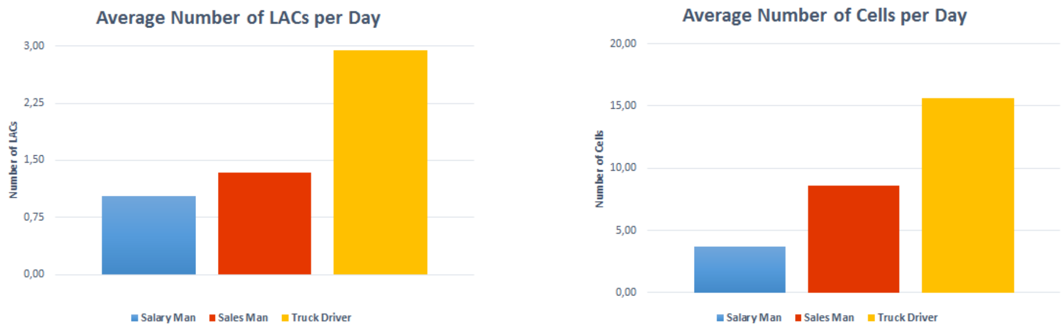


(d) Access Time

Figure 4.10: Sales Man and Truck Driver Access Results

4.4.3 LAC AND CELL COVERAGE

The number of LAN and Cell changes was also considered as presented in Figure 4.11, Figure 4.11a and Figure 4.11b indicate that the *Truck Driver* is the most mobile agent with larger numbers of LACs and Cells changed; it is also indicated that the *Salary Man* is the least mobile agent. Again, the results follow the agent's characteristics.



(a) LAC Changes

(b) Cell Changes

Figure 4.11: LAC and Cell changes

4.5 CHAPTER CONSIDERATIONS

This chapter detailed the simulation environment that was used as the base of implementation for the work presented in this thesis, along with existing results. This environment was initially developed with the goal of evaluating the deployment of ANDSF in an operator network covering a city populated with agents of three kinds of profiles each with different daily routines. Simulation results with metrics such as number of signaling events (policy requests from the client), policy data exchanged and network access type were evaluated. It was noted that this simulation does not evaluate 3GPP cell congestion, and as the default policy is to perform data offload to the available Wi-Fi AP, users may face bad QoE as the HotSpot becomes congested. Moreover, no evaluation is performed regarding the QoS of the Wi-Fi service. Therefore, the simulation should be extended to evaluate a) the impact of the ANDSF policy decision tree in terms of AP congestion, b) the congestion level at the mobile Cell in the mall, and c) the impact of different policies, that take cell congestion into account, on congestion of the SSIDs in the mall and on the generated signaling.

CHAPTER 5

Congestion Management with Enhanced ANDSF

This chapter will detail the core work of this dissertation, both the network architecture and the simulation environment envisioned to enhance the ANDSF-S with cell congestion information. In this way, the issues raised in the previous chapter (section 4.5) are addressed: i) perform an evaluation on the impact of the ANDSF policy decision tree in terms of AP congestion, ii) assess the congestion level at the mobile Cell in the mall, and iii) determine the impact of different policy decision trees, that take cell congestion into account, on the congestion level of the SSIDs in the mall and on the generated signaling. The cell congestion information will reach the ANDSF rules engine through the PCRF, provided via the UPCON mechanism.

5.1 Introduction

In the recently frozen Release 13 of the PCC and ANDSF standards ([28] and [30]), an interface between the ANDSF Server and the PCRF in not yet defined, therefore rules are defined according to time of day, location and user profile, while not taking into account network conditions (i.e are defined statically). Nevertheless, that interface exists in commercial products [60].

As detailed in Chapter 3 (section 3.2), with UPCON the PCRF receives congestion information from the RCAF and is able to use that data to send policies to CN entities in order to mitigate congestion in the user-plane. The standardized solution focuses this mitigation on traffic prioritization, reduction and limitation of traffic based on user's subscription, application, and type of content at a CN level. But once that information is processed by the PCRF, it may be used for creating dynamic rules that react to congestion conditions at the 3G/LTE cell.

5.2 PROPOSED ARCHITECTURE AND USE CASES

The proposed solution requires the deployment of the ANDSF and UPCON mechanisms by the MNO following the architecture presented in Figure 5.1, where the relevant elements of the network are highlighted. The UE (with an ANDSF-Client) maintains a connection via the S14 interface to the ANDSF-Server for policy requests and provision, as in the architecture described in section 4.2, with the added functionality for Cell Congestion provided by the RCAF node that enables the RUCI reporting to the PCRF.
For this work, all the simulation elements (maps, agent profiles and user triggers) described in section 4.3 were reused. Taking into consideration that the only areas defined for agent gathering are Malls (at lunch time), it is assumed for this work that only Mall cells and APs may face congestion. This emulates a common scenario where a busy mall has the potential to have its AP cells overloaded, as people bring different devices and connect them at the same time (personal laptop, smartphone, smart-watch, etc.). It is also considered that the QoS provided by the cell is superior to that of the Wi-Fi APs, since QoS enabled Wi-Fi is not yet common.

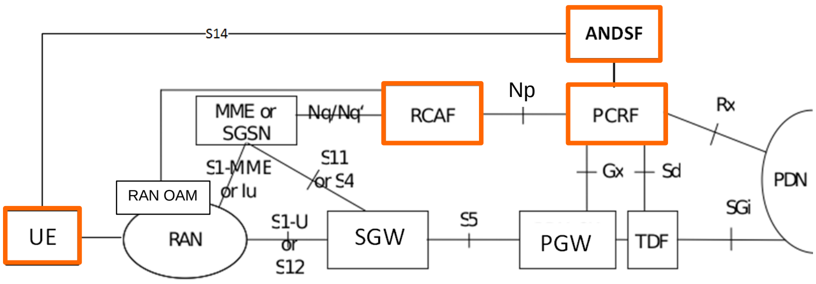


Figure 5.1: Implemented network architecture

5.2.1 USE CASES AND POLICIES

Two main Use Cases were defined to evaluate the proposed solution regarding the improvement of Wi-Fi QoS and for each case four different policies were analyzed. In each Case, the mobile Cell in the mall is designed to support service at full capacity to a number equal to half the total amount of agents in the simulation. Congestion reports are triggered on every congestion status change, corresponding to the scenarios described in subsection 3.3.1 (an user enters a cell) and subsection 3.3.2 (an user leaves a cell). It must be noted that the Premium SSIDs only serve premium users.

* *Case_A*: the SSIDs in the malls are designed to support service, at full capacity, to a number equal to the amount of agents in each map of the simulation.

* *Case_B*: the SSIDs in the malls are designed to support service, at full capacity, to a number equal to half the total amount of agents in each map of the simulation.

The Use Cases were selected to assess two different network deployments, in a mall, where Wi-Fi provides traffic offloading to prevent congestion at the mobile Cell. *Case_A* corresponds to an extreme and costly deployment where the Wi-Fi network is able to support the equivalent to all the population in a city. While the deployment in *Case_B* only supports the equivalent to half of that population and was selected because, in preliminary results, it indicated a significant level of congestion when

Policy_1 was applied.
The defined Policies which correspond to different ANDSF policy decision trees are now presented:

1. *Policy_1* - Send all agents to Wi-Fi: this policy corresponds to the ANDSF decision tree (in Figure 4.6) as implemented in subsection 4.3.5 and was used to set the baseline values for AP congestion periods (dubbed Bad_Wi-Fi periods);
2. *Policy_2* - Send all agents to Wi-Fi only when the Cell is congested: the policy decision flow (in Figure 5.2) takes into account mobile Cell congestion information. According to the policy, agents should remain in 3G while the mobile Cell is not congested;

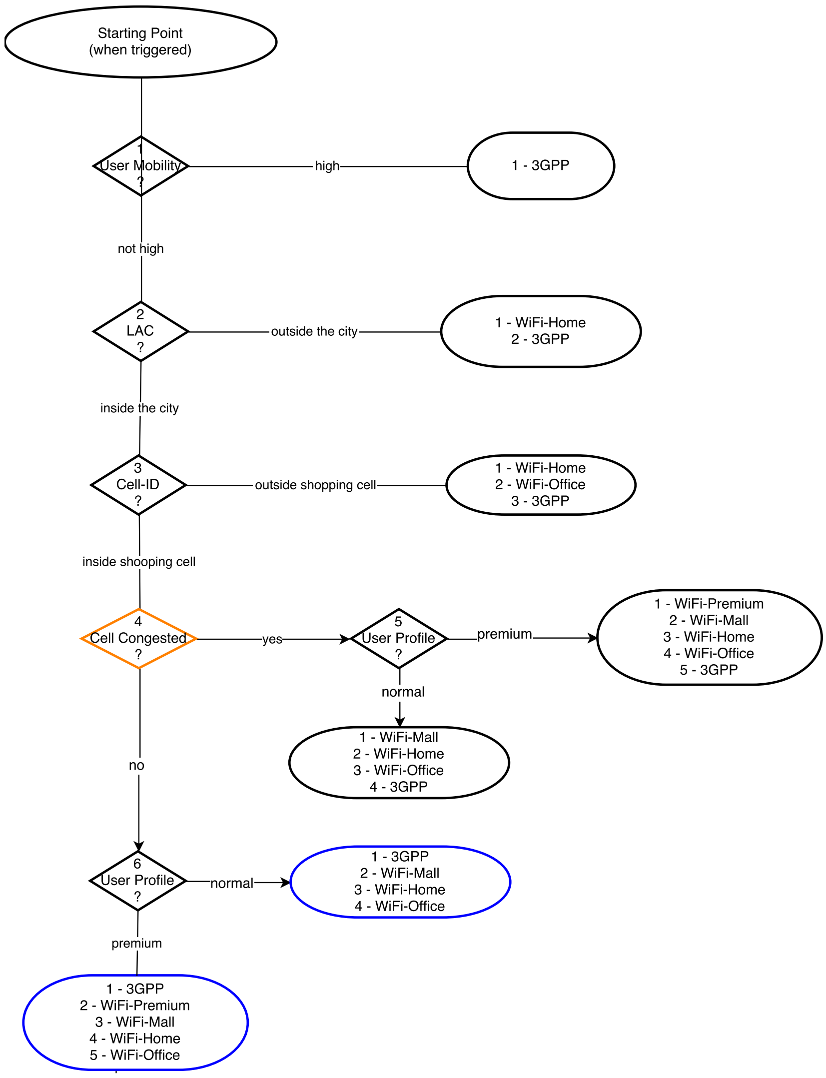


Figure 5.2: *Policy_2* decision tree

3. *Policy_3* - Send all non-Premium agents to Wi-Fi when the mobile Cell is congested: a policy where Premium agents remain connected to the mobile network when Cell congestion is detected (Figure 5.3). As all other agents are offloaded to Wi-Fi, congestion in the mobile Cell is always mitigated;

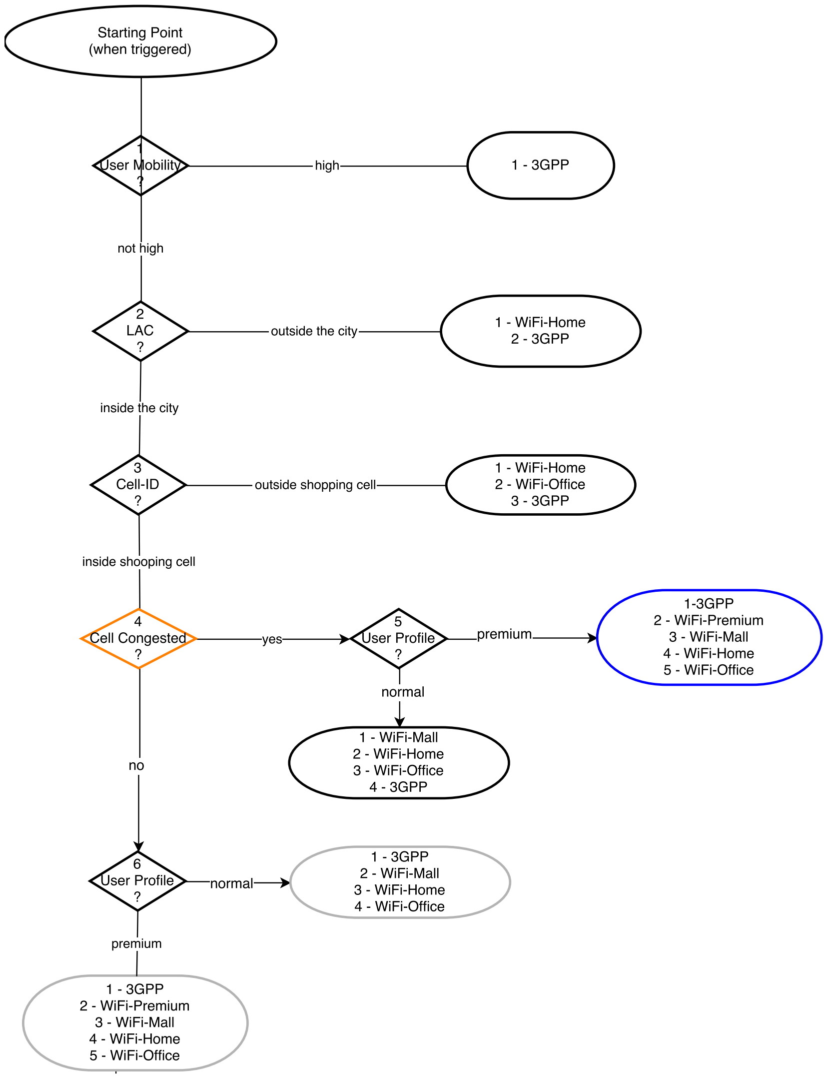


Figure 5.3: *Policy_3* decision tree

4. *Policy_4* - Send 80% of non-Premium agents to Wi-Fi when the Cell is congested: load balance between the mobile Cell and the Wi-Fi APs is further explored implementing a policy (in Figure 5.4) where Premium agents and 20% of non-Premium agents (randomly selected) are always kept connected to the mobile network. Again, as 80% of non-Premium agents are offloaded to Wi-Fi, cell congestion is always mitigated.

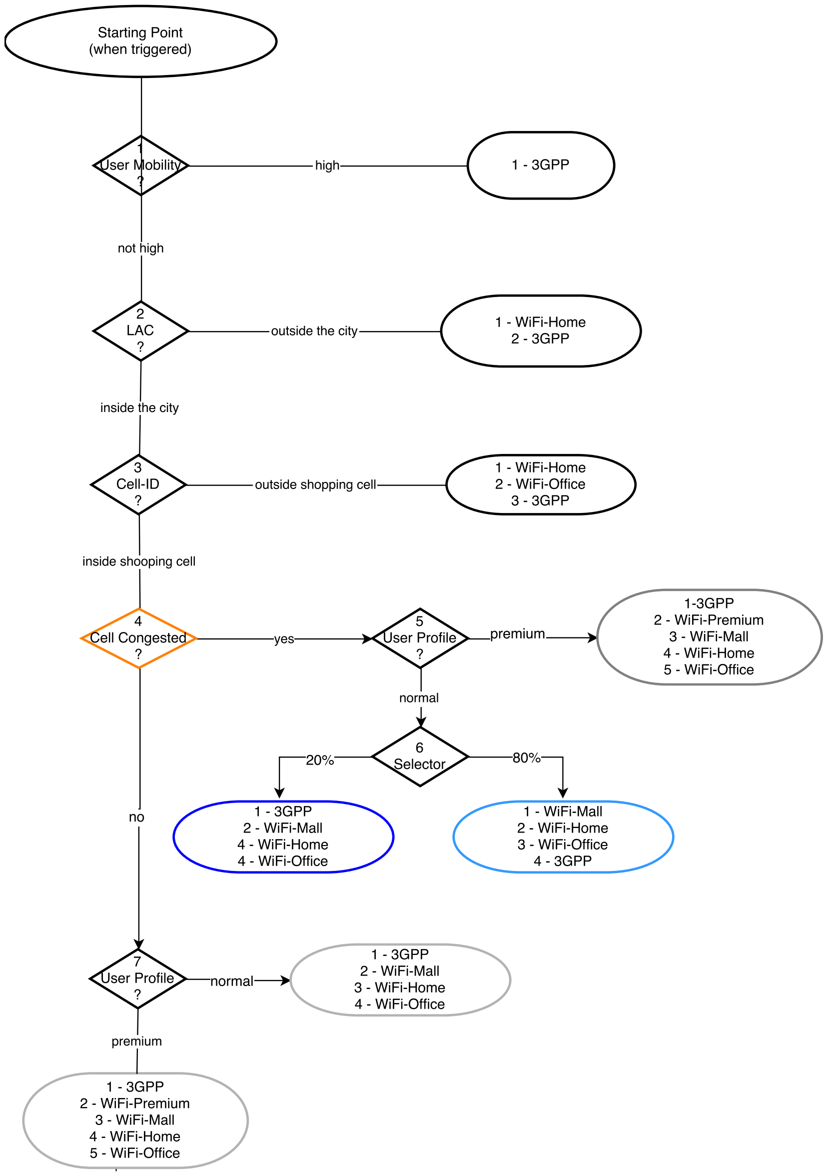


Figure 5.4: *Policy_4* decision tree

5.3 SIMULATION IMPLEMENTATION

Building on top of the existing simulation maintaining triggers, agent profiles, places and use-cases and addressing the issues raised in the previous chapter, the following features were implemented:

- Mall Wi-Fi APs congestion monitor: measures in real-time the congestion of the Mall and Premium APs and returns the time period when the agent is experiencing congestion (*Bad_Wi-Fi*).
- Mall mobile Cell congestion monitor: measures in real-time the congestion in the Mall cell and triggers congestion reporting via UPCON.

- Congestion event Trigger (*Trigger_14*): when congestion status changes in the Mall, cell policies are sent from the server to the client.
- Congestion Policies: different policies and corresponding scenarios (defined in subsection 5.2.1) were implemented that take cell congestion into account.

As mentioned in section 4.2, the Mall and Premium APs are aggregated under a single SSID.

The Cell Congestion Monitor algorithm uses a simulator API that returns the total number of agents in a specified range, corresponding to the area representing either a cell or an SSID. When the number of agents covered is higher than the congestion threshold, *Trigger_14* is activated. A congestion reporting procedure via UPCON signaling is initiated and updated policies are provided to the agents. As the mobile Cell covering the mall is composed of several sub-cells, it is assumed in this simulation that its overall cover range is the same as the Mall SSID.

The algorithm implementing the congestion monitor is similar to the Cell Congestion Monitor but takes in consideration that two different SSIDs are located in the mall:

- Mall SSID: used by normal users and by Premium users when not in range of the Premium SSID;
- Premium SSID: used only by Premium users.

5.3.1 SIGNALING

The implemented solution has two basic modes of operation in terms of signaling according to the Cell congestion status as presented in Figure 5.5. In the first, when no change of congestion is detected, the standard mode of operation of ANDSF, already presented in the previous chapter, is performed. When the UE notices a change in conditions, it initiates an ANDSF Pull Mode policy provision procedure (Figure 5.5-1) which is then answered by the ANDSF server (Figure 5.5-2) with the appropriate policy. The second is performed when congestion conditions change at the Cell and is initiated with a congestion notification to the RCAF (Figure 5.5-3), which triggers the UPCON congestion reporting procedure (Figure 5.5-4 to 7). The PCRF then determines the appropriate rules (according to the policy defined in subsection 5.2.1) which are sent to the Client via an ANDSF Push Mode policy distribution procedure (Figure 5.5-8 and 9).

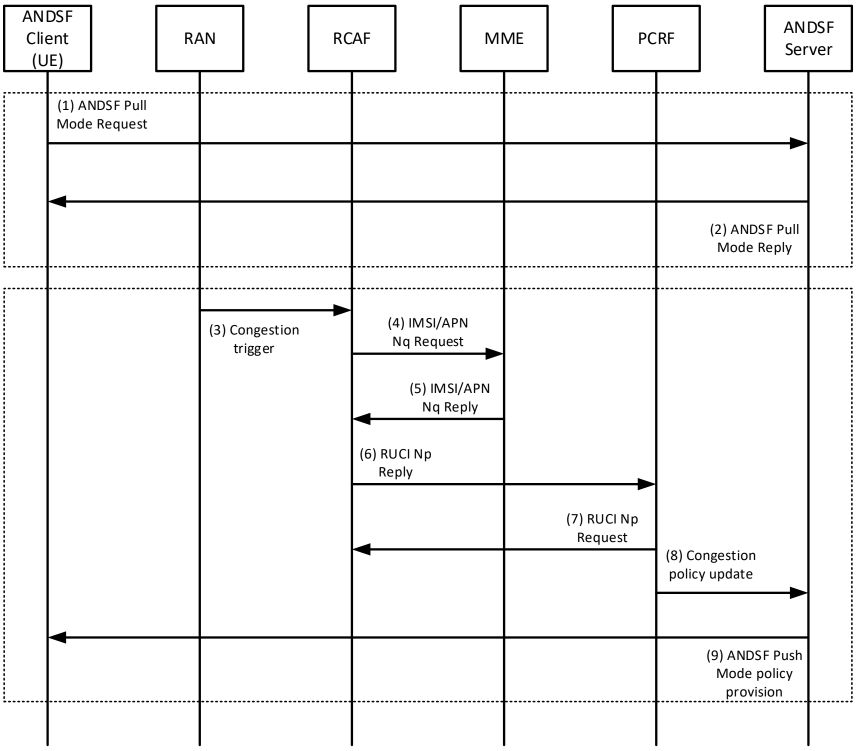


Figure 5.5: Signaling flow for the implemented architecture

In order to get results on the signaling impact of the RUCI report procedure on the MNO's network it was necessary to estimate the amount in bytes of the messages involved. Analyzing the messages of the Nq, Np protocols described in section 3.4 and the Diameter AVPs in [61] it was possible to verify the variability of the message sizes according to different conditions. For practicality, it was assumed that messages are exchanged without errors or optional AVPs, the RCAF sends information requests only to the MME (NqAP-IMSI-APN-INFORMATION-REQUEST and NqAP-IMSI-APN-INFORMATION-RESPONSE messages), RUCI reports are of the type Aggregated (ARR and ARA messages) and the variable information containing the number of agents that caused congestion should be considered. The estimated size of the Nq and Np messages is presented in Tables 5.1 and 5.2 respectively.

Table 5.1: Estimated size of Nq messages

NqAP-IMSI-APN-INFORMATION-REQUEST		NqAP-IMSI-APN-INFORMATION-RESPONSE	
Information Element	Bytes	Information Element	Bytes
Message type	1	Message type	1
RAN Entity Identifier	11	Cause	4
		RAN Associated Information	14 + 22 * #UE's
		MME Name	7
Size =	12	Size =	26 + 22 * #UE's

Table 5.2: Estimated size of Np messages

<AR-Request >		<AR-Answer >	
AVP	Bytes	AVP	Bytes
Code	4	Code	4
Flags	1	Flags	1
Length	3	Length	3
Vendor ID	4	Vendor ID	4
<Session-Id >	4	<Session-Id >	4
{ Vendor-Specific-Application-Id }	4	{ Vendor-Specific-Application-Id }	4
{ Auth-Session-State }	4	{ Auth-Session-State }	4
{ Origin-Host }	16	{ Origin-Host }	16
{ Origin-Realm }	16	{ Origin-Realm }	16
{ Destination-Realm }	16		
Aggregated-RUCI-Report	8		
{ Aggregated-Congestion-Info }	8		
[Congestion-Location-Id]	19		
[IMSI-List]	8 + 8 * #IMSIs		
[Congestion-Level-Value]	12		
+ Diameter header	20	+ Diameter header	20
Size =	137 + 8 * #IMSIs	Size =	72

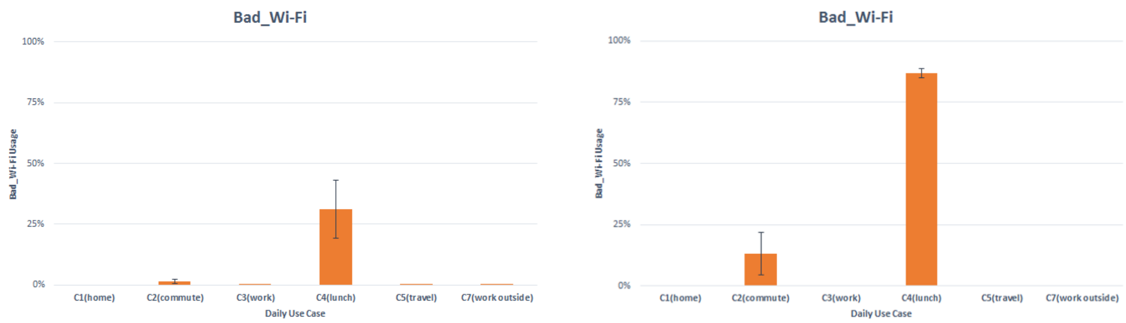
5.4 EVALUATION

The procedure to perform the evaluation of this solution was identical the one described in section 4.4, running three times for the various policies in each Use Case. In addition to the metrics analyzed in that section, the measures on *Bad_Wi-Fi*, corresponding to the periods of poor Wi-Fi service and the signaling impact of the RUCI reporting procedure, were considered. Average results with 95% of confidence are presented in the following sections.

5.4.1 BAD_WI-FI

In this section, the simulation results regarding the *Bad_Wi-Fi* metric will be presented. Results for the application of *Policy_1*, are presented in Figures 5.6, 5.7 and 5.8. The simulation results for the application of *Policy_2* are presented in Figures 5.9, 5.10 and 5.11. For *Policy_3*, the results are presented in Figures 5.12, 5.13 and 5.14. And in Figures 5.15, 5.16 and 5.17 the results for the application of *Policy_4* are presented. A summary of the results is presented in Table 5.3.

Figure 5.6 presents the average periods of congested Wi-Fi usage (*Bad_Wi-Fi*) when compared with the total Wi-Fi coverage, in each of the various Cases defined in subsection 4.3.2 that represent the whereabouts of the agents along the day, when *Policy_1* is applied. Analyzing Figure 5.6a, which presents the *Case_A*, it is evident that the time period with the most significant portion of congested Wi-Fi usage is lunch time (on average 31.10 ± 11.97% in C4). The time of *Bad_Wi-Fi* in other Cases is at most three seconds. This is justified by the fact that all agents gather in the mall in their lunch period. From Figure 5.6b, with *Case_B*, it is also evident that the largest period of congested Wi-Fi usage is lunch time (C4) with, on average, 86.84 ± 1.94% of the total Wi-Fi coverage. The period of *Bad_Wi-Fi* in other Cases it at most 52 seconds in the work period (C3). Comparing the results for *Case_A* and *Case_B* it is evident the effect of deploying Wi-Fi SSIDs at the Mall that support only half the agents in the simulation. In fact, at lunch time, the time of *Bad_Wi-Fi* in *Case_A* corresponds, on average, to 31.10% of the time while in *Case_B* corresponds to 86.84% of the Wi-Fi coverage time.



(a) *Case_A* (b) *Case_B*

Figure 5.6: *Bad_Wi-Fi* with *Policy_1*

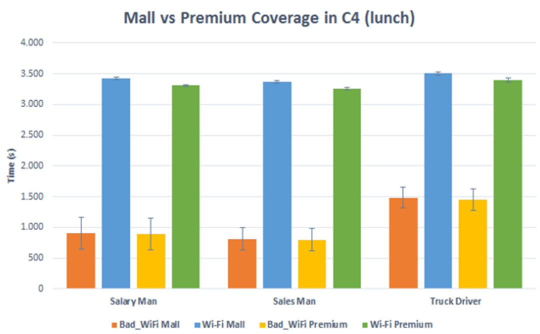
In Figure 5.7 the lunch time period (C4) is analyzed, presenting the average period, in seconds, of total Wi-Fi coverage (*Wi-Fi*) and congested Wi-Fi usage (*Bad_Wi-Fi*) for each of the three agent profiles defined in subsection 4.3.2. From Figure 5.7a, where the *Case_A* is presented, the results show that the periods of poor service represent 26.41 ± 7.53%, 24.16 ± 5.41% and 42.38 ± 4.89% of the total Wi-Fi service for the *Salary Man*, the *Sales Man* and the *Truck Driver*, respectively. This discrepancy between the values for *Bad_Wi-Fi* among the *Truck Driver* and other agents is justified by the fact that this agent is very mobile and other agents only face congestion when it reaches the mall. Regarding *Case_B*, Figure 5.7b shows that the *Truck Driver* uses, on average, around 100 seconds more Wi-Fi and *Bad_Wi-Fi* than the other agents. Furthermore, the periods for poor Wi-Fi service in *Case_B* with *Policy_1* are almost of the same magnitude. In fact, comparing the values for each type of agent, the difference between the total coverage and poor service is, on average, around 450 seconds. The period of *Bad_Wi-Fi* corresponds to around 87% of the total Wi-Fi coverage. Comparing both Cases, the difference between the periods of total coverage are minimal, but regarding the poor Wi-Fi, the values double for the *Truck Driver* and more than triple for the *Salary Man* and *Sales Man* agents. **The impact of the reduced capacity of the SSIDs in the Mall is evident when *Policy_1* is applied.**

(a) *Case_A* (b) *Case_B*

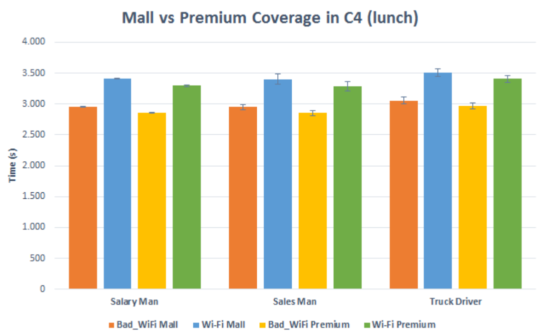
Figure 5.7: *Bad_Wi-Fi* in C4 (lunch) with *Policy_1*

In Figure 5.8, the lunch time period (C4) is further analyzed, presenting the average period, in seconds, of total Wi-Fi coverage at the Mall and Premium SSIDs (*Wi-Fi Mall* and *Wi-Fi Premium* respectively), and congested Wi-Fi usage at the the Mall and Premium SSIDs (*Bad_Wi-Fi Mall* and

Bad_ Wi-Fi Premium respectively) for each of the three agent profiles defined in subsection 4.3.2. From Figure 5.8a, where *Case_A* is presented, the results follow those shown in Figure 5.7, with the usage time of *Bad_ Wi-Fi Mall* being 26.41 ± 7.53%, 24.16 ± 5.41% and 42.38 ± 4.89% of the total service at the Mall SSID for each agent, respectively. The discrepancy between the three agents is justified by the agent mobility as stated before. These are the same results as in Figure 5.7a as the periods referred there are the maximum values for poor service which correspond to *Bad_ Wi-Fi Mall* periods. Regarding the Premium SSID, the time period for poor QoS represents 26.90 ± 7.80%, 24.58 ± 5.60% and 42.78 ± 5.04% of the total Wi-Fi coverage for each agent respectively. For *Case_B*, in Figure 5.8b, the results follow those presented in Figure 5.7b for the *Bad_ Wi-Fi Mall*. The periods of Premium coverage are, on average, 100 seconds inferior to those on the Mall SSID, and relatively, the poor service represents around 87% of the total. Comparing *Case_A* and *Case_B*, the length of the poor Wi-Fi service rises from around 25% in the case of the *Salary Man* and *Sales Man* or 42% for the *Truck Driver* to around 87% in all of them, again, showing the impact of the reduced coverage capacity of the SSIDs in the Mall when this policy is applied.



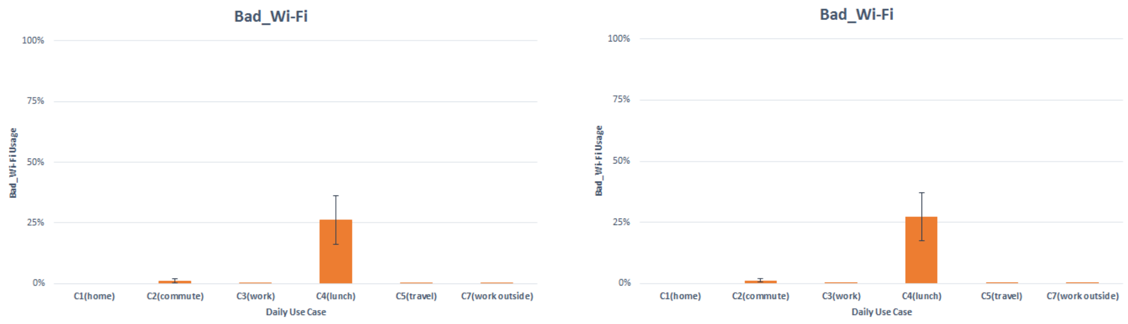
(a) *Case_A*



(b) *Case_B*

Figure 5.8: Mall vs. Premium *Bad_ Wi-Fi* with *Policy_1*

In Figure 5.9, which presents the average periods of congested Wi-Fi usage (*Bad_ Wi-Fi*) when compared with the total Wi-Fi coverage, in each of the various Cases representing the whereabouts of the agents along the day, when *Policy_2* is applied. Analyzing 5.9a with the results for *Case_A*, it is clear that the time period with the most significant usage of poor Wi-Fi service is the lunch time (C4) with 26.17 ± 10.05% of the total Wi-Fi coverage. The time of *Bad_ Wi-Fi* in other daily periods is at most 3 seconds in work (C3). Regarding *Case_B*, in Figure 5.9b, it is also clear that it is during lunch time (C4) that the usage of poor Wi-Fi service is more significant with 27.27 ± 9.90% of the total coverage. Comparing *Case_A* and *Case_B* is evident that the periods of *Bad_ Wi-Fi* are now almost identical, being slightly larger in *Case_B*. This contrasts with the results presented in Figure 5.6 for *Policy_1*, in *Case_A* the period of *Bad_ Wi-Fi* during lunch time (C4) was reduced, on average, from around 31% to 26%, but most significantly, in *Case_B* the period for poor QoS was reduced from 87% to 27%. This result implies that the mall mobile Cell is not congested for long periods during lunch time. **Therefore, with this policy, the improvement in QoS is very significant in *Case_B*.**



(a) *Case_A* (b) *Case_B*

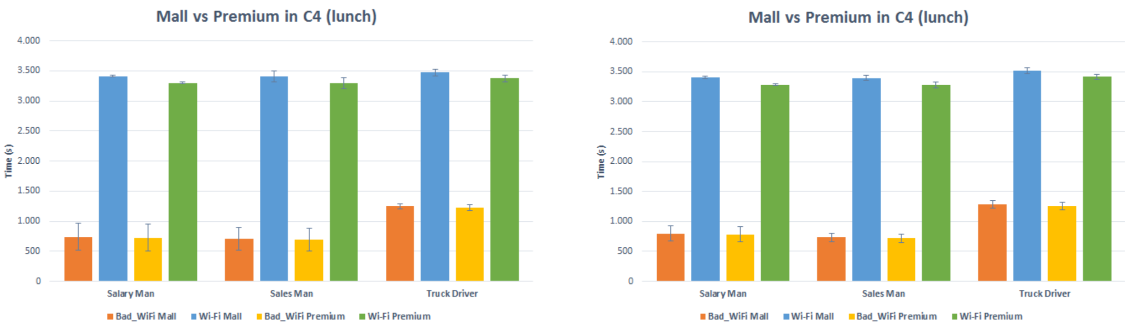
Figure 5.9: *Bad_Wi-Fi* with *Policy_2*

Figure 5.10 presents an analysis of the lunch time period (C4) when *Policy_2* is applied, with the average period, in seconds, of total Wi-Fi coverage (*Wi-Fi*) and congested Wi-Fi usage (*Bad_Wi-Fi*) for each of the three agent profiles. In Figure 5.10a, for *Case_A*, the results show that the periods of poor Wi-Fi service represent 21.70 ± 6.57%, 20.66 ± 5.44% and 35.96 ± 1.25% of the total periods of Wi-Fi coverage for the *Salary Man*, *Sales Man* and *Truck Driver* agents, respectively, being the difference justified, again, by the different mobility of the agent. In *Case_B* (Figure 5.10b), the periods of poor QoS are 22.08 ± 6.80%, 21.01 ± 5.64% and 36.01 ± 1.29% for the three types of agents, respectively. These results follow those presented in Figure 5.9 where the periods for poor service are similar in *Case_A* and *Case_B*. Comparing with Figure 5.7 which presents the results for *Policy_1*, applying this policy in *Case_A* reduces the periods of poor service, on average, from around 26%, 24% and 42% to 22%, 21% and 36% for the *Salary Man*, *Sales Man* and *Truck Driver* agents respectively. For *Case_B* the *Bad_Wi-Fi* is reduced from around 87% to 23%, 22% and 37% for the three types of agents respectively. **The improvement in QoS with the application of *Policy_2* exists in both Cases and is most noticeable in *Case_B*.**

(a) *Case_A* (b) *Case_B*

Figure 5.10: *Bad_Wi-Fi* in C4 (lunch) with *Policy_2*

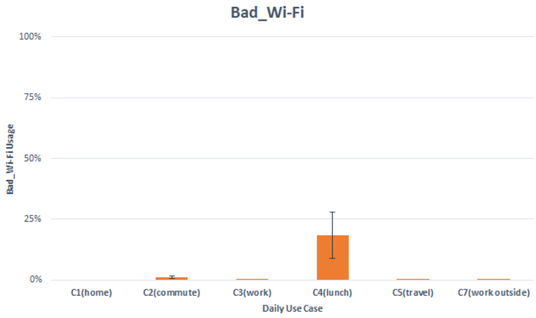
In Figure 5.11 the lunch time period (C4) when *Policy_2* is applied is further analyzed, presenting the average period, in seconds, of total Wi-Fi coverage at the Mall and Premium SSIDs (*Wi-Fi Mall* and *Wi-Fi Premium* respectively), and congested Wi-Fi usage at the the Mall and Premium SSIDs (*Bad_Wi-Fi Mall* and *Bad_Wi-Fi Premium* respectively) for each of the three agent profiles defined in subsection 4.3.2. From Figure 5.11a where the *Case_A* is presented, the results for *Bad_Wi-Fi Mall* correspond to 21.70 ± 6.57%, 20.66 ± 5.44% and 35.96 ± 1.25% of the periods of total Wi-Fi coverage at the Mall SSID for the *Salary Man*, *Sales Man* and *Truck Driver* agents respectively. Regarding the poor Premium service, the results represent 22.08 ± 6.80%, 21.01 ± 5.64% and 36.01 ± 1.29% of total coverage periods in the Premium SSID. For *Case_B*, in Figure 5.11b, the results for poor Wi-Fi service at the Mall SSID correspond to 23.46 ± 3.66%, 21.55 ± 2.01% and 36.47 ± 1.69% of the periods for total Wi-Fi coverage for each agent type, respectively. Regarding the poor Premium Wi-Fi service periods, their length correspond to 23.90 ± 3.79%, 21.92 ± 2.08% and 36.81 ± 1.74% of the periods of total coverage for each agent type, respectively. Comparing both Cases, the periods for poor QoS in either SSIDs is nearly identical, albeit higher in *Case_B*. Comparing these results with Figure 5.8 which presents the results for *Policy_1*, with this policy, in *Case_A*, the bad Wi-Fi service periods were reduced, on average, by around 5%, 3% and 6% for the three different agents respectively. But for *Case_B* the reduction is of around 87% to 24%, 22% and 37%.



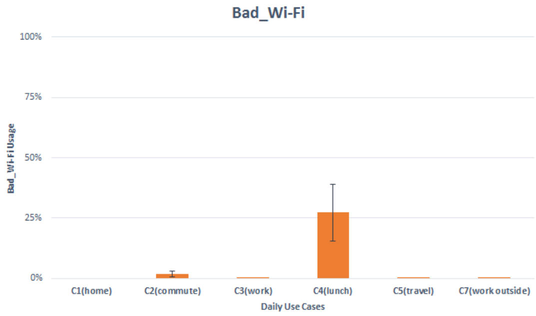
(a) *Case_A* (b) *Case_B*

Figure 5.11: Mall vs. Premium *Bad_Wi-Fi* with *Policy_2*

Figure 5.12 presents the average periods of congested Wi-Fi usage (*Bad_Wi-Fi*) when compared with the total Wi-Fi coverage, in each of the various cases representing the whereabouts of the agents along the day, when *Policy_3* is applied. Performing an analysis on Figure 5.12a, for *Case_A*, it is evident that the most significant usage of poor Wi-Fi service occurs during lunch time with 18.47 ± 9.35% of the total Wi-Fi coverage. In other periods of the day, the usage of *Bad_Wi-Fi* is at most four seconds during work (C3). For *Case_B*, in Figure 5.12b, the usage of service with poor QoS happens during the lunch period (C4) with 27.32 ± 11.74% of the total Wi-Fi coverage. Comparing *Case_A* with *Case_B*, the usage of congested Wi-Fi is more significant during lunch time (C4) and is around 9% longer in *Case_B*. Performing a comparison with the results for the application of *Policy_2* (in Figure 5.9), the periods of *Bad_Wi-Fi* during lunch (C4) were reduced, on average, from around 26% to 18% in *Case_A* and is slightly larger in *Case_B*. Comparing with the results in Figure 5.6, for *Policy_1*, the usage of poor Wi-Fi is reduced from around 30% to 18% and, most significantly, from 87% to 27% for *Case_A* and *Case_B* respectively. **The results imply that *Policy_3* is at least as good as *Policy_2* on improving the QoS and much better than *Policy_1*.**



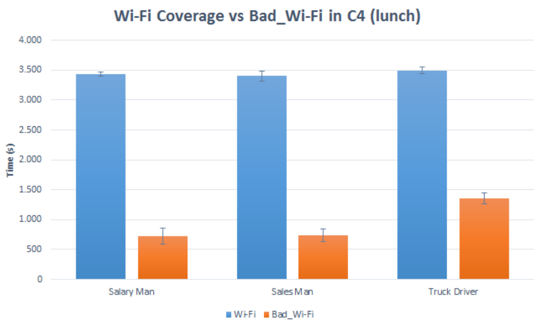
(a) *Case_A*



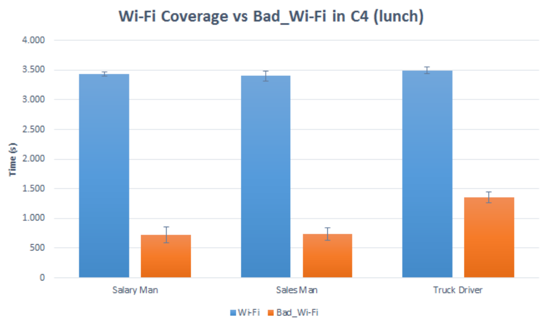
(b) *Case_B*

Figure 5.12: *Bad_Wi-Fi* with *Policy_3*

Figure 5.13 presents an analysis of the lunch time period (C4) when *Policy_3* is applied, with the average period, in seconds, of total Wi-Fi coverage (*Wi-Fi*) and congested Wi-Fi usage (*Bad_Wi-Fi*) for each of the three agent profiles. In Figure 5.13a, for *Case_A*, the periods for poor QoS represent 13.81 ± 2.05%, 13.81 ± 1.72% and 27.59 ± 4.71% of the values for total Wi-Fi coverage for the *Salary Man*, *Sales Man* and *Truck Driver* agents respectively. For *Case_B*, in (Figure 5.13b) the results show periods of *Bad_Wi-Fi* which correspond to 21.23 ± 3.87%, 21.77 ± 3.09% and 38.68 ± 2.52% versus the periods for total coverage for each of user types, respectively. Comparing both Cases, the periods for poor Wi-Fi service for each agent type are significantly higher in *Case_B* (around 7%, 8% and 11% on average for each type of user respectively). Performing a comparison with *Policy_2* (in Figure 5.10), in *Case_A* the poor Wi-Fi periods were reduced by around 8%, 7% and 8%, on average, for the three types of agents respectively, while in *Case_B* the poor QoS was reduced for the *Salary Man* and increased slightly for the *Sales Man* and *Truck Driver* agents. Comparing with Figure 5.7, which presents the results for *Policy_1*, with the application of *Policy_3* in *Case_A*, the periods of poor service are reduced from around 26%, 24% and 42% to 14%, 14% and 28% for the Salary Man, Sales Man and Truck Driver agents respectively. Regarding *Case_B*, the poor QoS periods were reduced, on average, from around 87% to 21%, 22% and 39% for the three types of agents respectively. **Therefore, *Policy_3* performs a clear improvement in QoS versus *Policy_1*, which is more noticeable in *Case_B*, and achieves better results than *Policy_2* in *Case_A*.**



(a) *Case_A*



(b) *Case_B*

Figure 5.13: *Bad_Wi-Fi* in C4 (lunch) with *Policy_3*

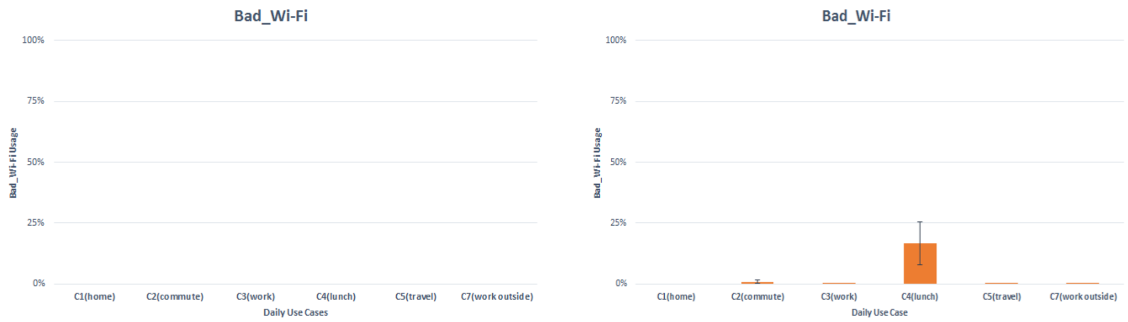
A further analysis of the lunch time period (C4), when *Policy_3* is applied, is presented in

Figure 5.14, with the average period, in seconds, of total Wi-Fi coverage at the Mall and Premium SSIDs (*Wi-Fi Mall* and *Wi-Fi Premium* respectively), and congested Wi-Fi usage at the the Mall and Premium SSIDs (*Bad_Wi-Fi Mall* and *Bad_Wi-Fi Premium* respectively) for each of the three agent profiles defined in subsection 4.3.2. From Figure 5.14a for *Case_A*, the results for *Bad_Wi-Fi Mall* correspond to 13.81 ± 2.05%, 13.81 ± 1.72% and 27.59 ± 4.71% of the periods of the total coverage at the Mall SSID for the *Salary Man*, *Sales Man* and *Truck Driver* users, respectively. The periods of poor Premium service are non-existent as all Premium agents are kept connected to the mobile Cell. For *Case_B*, in Figure 5.14b, the results for poor Wi-Fi service at the Mall SSID correspond to 21.23 ± 3.87%, 21.77 ± 3.09% and 38.68 ± 2.52% of the periods of the total coverage for each agent type respectively. For poor Premium Wi-Fi service, no periods are also verified as Premium agents do not use Wi-Fi. Comparing both Cases, the periods for poor QoS in either SSIDs are almost equivalent, being slightly larger in *Case_B*. Performing a comparison with Figure 5.11 which presents the results for the application of *Policy_2*, in *Case_A*, the periods for poor QoS were reduced in the Mall SSID, on average, from around 22%, 21% and 36% to 14%, 14% and 28% and to zero in the Premium SSID for each of the types of agents, respectively. Regarding the *Case_B*, the periods of poor Wi-Fi service in the Mall SSID are reduced in around 2% for the *Salary Man* and slightly increased to the *Sales Man* and *Truck Driver*, respectively. For the Premium SSID the congested wi-Fi service is reduced to zero for all agents. Comparing these results with Figure 5.8 which presents the results for the application of *Policy_1*, with *Policy_3*, in *Case_A*, the bad Wi-Fi service periods in the Mall SSID were reduced, on average, by around 13%, 10% and 15% for the three different agents respectively. And for the Premium SSID, the poor service is reduced to zero. For *Case_B*, the reduction is of around 87% to 21%, 22% and 39% in the Mall SSID and to zero in the Premium SSID.

(a) *Case_A* (b) *Case_B*

Figure 5.14: Mall vs. Premium *Bad_Wi-Fi* with *Policy_3*

In Figure 5.12 a comparison is performed between the average periods of congested Wi-Fi usage (*Bad_Wi-Fi*) with the total Wi-Fi coverage, in each of the various cases representing the whereabouts of the agents along the day, when *Policy_4* is applied. Analyzing Figure 5.15a, for *Case_A*, it is evident that with the application of this policy the usage of *Bad_Wi-Fi* is zero in this Case. For *Case_B* (in Figure 5.15b), the period of the day where poor Wi-Fi is used is during lunch (C4) with 16.76 ± 8.84% of total Wi-Fi coverage. Comparing both results the most significant fact is that in *Case_A* there is no period of poor QoS. Performing a comparison with Figure 5.12 with the results for the application of *Policy_3*, for *Case_A* the period of poor Wi-Fi usage during lunch was reduced, on average, from around 18% to zero and from 27% to 18% in *Case_B*. Comparing with the results for the application of *Policy_2* (in Figure 5.9), the periods of poor QoS were reduced in *Case_A*, on average, from around 26% to zero and in *Case_B* from 27% to 18%. Comparing with the results for *Policy_1* (in Figure 5.6), the periods of poor QoS were reduced, on average, from around 31% to zero in *Case_A* and from 87% to 18% *Case_B*. **These results imply that *Policy_4* is superior to all the other three policies in reducing poor Wi-Fi QoS enabling zero congestion when the SSIDs support all the agents in the simulation.**



(a) *Case_A* (b) *Case_B*

Figure 5.15: *Bad_Wi-Fi* with *Policy_4*

Figure 5.16 presents an analysis of the lunch time period (C4) when *Policy_4* is applied, with the average period, in seconds, of total Wi-Fi coverage (*Wi-Fi*) and congested Wi-Fi usage (*Bad_Wi-Fi*) for each of the three agent profiles. From Figure 5.16a, regarding *Case_A*, it is clear that with this policy there are no periods of poor Wi-Fi service for any of the agent types. For *Case_B*, the poor QoS periods represent 12.72 ± 1.96%, 11.87 ± 1.97% and 25.61 ± 5.28% of the total Wi-Fi coverage. Comparing both Cases, in *Case_B* the poor Wi-Fi still exists and is more significative for the *Truck Driver* agent. Comparing with *Policy_3* (in Figure 5.13), in *Case_A* the poor QoS periods were reduced, on average, from around 14%, 14% and 28% to zero for the three types of agents respectively, while in *Case_B* the poor Wi-Fi service was reduced, on average, in around 8%, 8% and 11% for the three types of agents, respectively. Performing a comparison with *Policy_2* (in Figure 5.10), in *Case_A* the poor Wi-Fi periods were reduced from around 22%, 21% and 36% to zero for the *Salary Man*, *Sales Man* and *Truck Driver*, respectively, while in *Case_B*, the poor QoS periods were reduced from around 87% to 13%, 12% and 26% for the three types of agents, respectively.

(a) *Case_A* (b) *Case_B*

Figure 5.16: *Bad_Wi-Fi* in C4 (lunch) with *Policy_4*

In Figure 5.17 the lunch time period (C4) when *Policy_4* is applied is further analyzed, presenting the average period, in seconds, of total Wi-Fi coverage at the Mall and Premium SSIDs (*Wi-Fi Mall* and *Wi-Fi Premium* respectively), and congested Wi-Fi usage at the the Mall and Premium SSIDs (*Bad_ Wi-Fi Mall* and *Bad_ Wi-Fi Premium* respectively) for each of the three agent profiles. In Figure 5.17a, for *Case_A*, the periods of poor Wi-Fi service at Mall and Premium SSIDs are null. For *Case_B*, in Figure 5.17b, the periods for poor Wi-Fi service correspond to 12.72 ± 1.96%, 11.87 ± 1.97% and 25.61 ± 5.28% of the total Wi-Fi service for each agent type respectively. Regarding the poor Premium Wi-Fi service, the periods are non-existent as the agents are kept connected to the mobile Cell. Comparing *Case_A* and *Case_B*, congestion periods are non-existing in the first Case but still remain in the second Case being, on average, twice as long for the *Truck Driver*. Performing a comparison with *Policy_3* (in Figure 5.13), in *Case_A* the poor QoS periods were reduced to zero from around 14% for the *Salary Man* and *Sales Man* agents and from 28% for the *Truck Driver*. In *Case_B* the poor Wi-Fi service at the Mall SSID was reduced in around 8%, 10% and 13% for the three types of agents respectively. At the Premium SSID no poor Wi-Fi service was verified with both policies. Performing a comparison with *Policy_2* (in Figure 5.10), in *Case_A* the poor Wi-Fi periods were reduced from around 22%, 21% and 36% to zero for the three types of agents respectively, while in *Case_B*, the poor QoS periods at the Mall SSID were reduced, on average, from around 87% to 13%, 12% and 26%, and to zero in the Premium SSID for the three types of agents, respectively.

(a) *Case_A* (b) *Case_B*

Figure 5.17: Mall vs. Premium *Bad_ Wi-Fi* with *Policy_4*

Comparing the results for the four different policies, it is evident that the application of Policies 2, 3 and 4, enabled by the existence of a congestion detection and reporting mechanism implemented in this solution, provide a significant improvement in the QoS at the lunch period when compared with *Policy_1* where such mechanism is not considered. This improvement is specially noticeable in *Case_B* with a decrease, on average, of *Bad_ Wi-Fi* from around 87% in both SSIDs to 27% in *Policy_2* and up to 17% in *Policy_4*. A discrepancy was detected between the interval of poor Wi-Fi service of the *Truck Driver* and other agents justified by the fact that this agent is very mobile and other types of agent only face congestion when it reaches the mall. Also relevant is that *Policy_4*, in *Case_A*, enabled the non-existence of bad QoS periods in both SSIDs. Moreover, with Policies 3 and 4 the Premium agents do not suffer any poor Wi-Fi service. These significant improvements in QoS for both SSIDs at the mall show the importance of the deployment of the solution proposed in this work. Regarding the decision on which Use Case to deploy, i.e, if the Wi-Fi SSIDs at the mall should support all or only half the agents in the simulation (*Case_A* and *Case_B* respectively), the MNO should deploy the first Case with *Policy_4* if it wants to guarantee no periods of poor QoS for premium and normal users. Deploying *Case_B* with *Policy_4* will cause some congestion for premium and normal users but for 50% of the deployment cost.

Table 5.3 presents a summary of the results for the *Bad_Wi-Fi* metric in the lunch period (C4). In each Use Case, for each Policy and agent type (*Salary Man*, *Sales Man* and *Truck Driver*) the relative percentages (and length, in seconds) of *Bad_Wi-Fi* versus the total Wi-Fi coverage periods in the Mall and Premium SSIDs are presented.

Table 5.3: *Bad_Wi-Fi* results in C4 (lunch) per Policy and Use Case

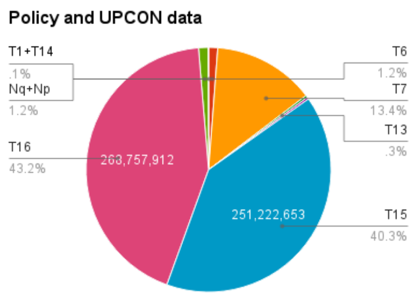
		Case_A		*Case_B*	
		Mall SSID	Premium SSID	Mall SSID	Premium SSID
Policy_1	Salary Man	26.41 ± 7.53% (905 ± 258 s)	26.90 ± 7.80% (890 ± 258 s)	86.64 ± 0.22% (2957 ± 8 s)	86.68 ± 0.23% (2855 ± 8 s)
	Sales Man	24.16 ± 5.41% (815 ± 182 s)	24.58 ± 5.60% (800 ± 182 s)	86.79 ± 1.22% (2954 ± 42 s)	86.84 ± 1.27% (2853 ± 42 s)
	Truck Driver	42.38 ± 4.89% (1485 ± 171 s)	42.78 ± 5.04% (1454 ± 171 s)	87.10 ± 1.47% (3057 ± 52 s)	87.15 ± 1.51% (2968 ± 52 s)
Policy_2	Salary Man	21.70 ± 6.57% (741 ± 224 s)	22.08 ± 6.80% (728 ± 224 s)	23.46 ± 3.66% (799 ± 125 s)	23.90 ± 3.79% (785 ± 125 s)
	Sales Man	20.66 ± 5.44% (704 ± 186 s)	21.01 ± 5.64% (691 ± 186 s)	21.55 ± 2.01% (732 ± 68 s)	21.92 ± 2.08% (719 ± 68 s)
	Truck Driver	35.96 ± 1.25% (1250 ± 44 s)	36.01 ± 1.29% (1225 ± 44 s)	36.47 ± 1.69% (1284 ± 60 s)	36.81 ± 1.74% (1257 ± 60 s)
Policy_3	Salary Man	13.81 ± 2.05% (474 ± 70 s)	0.00 ± 0.00% (0 ± 0 s)	21.23 ± 3.87% (728 ± 133 s)	0.00 ± 0.00% (0 ± 0 s)
	Sales Man	13.81 ± 1.72% (467 ± 58 s)	0.00 ± 0.00% (0 ± 0 s)	21.77 ± 3.09% (740 ± 105 s)	0.00 ± 0.00% (0 ± 0 s)
	Truck Driver	27.59 ± 4.71% (961 ± 164 s)	0.00 ± 0.00% (0 ± 0 s)	38.68 ± 2.52% (1353 ± 88 s)	0.00 ± 0.00% (0 ± 0 s)
Policy_4	Salary Man	0.00 ± 0.00% (0 ± 0 s)	0.00 ± 0.00% (0 ± 0 s)	12.72 ± 1.96% (437 ± 67 s)	0.00 ± 0.00% (0 ± 0 s)
	Sales Man	0.00 ± 0.00% (0 ± 0 s)	0.00 ± 0.00% (0 ± 0 s)	11.87 ± 1.97% (405 ± 67 s)	0.00 ± 0.00% (0 ± 0 s)
	Truck Driver	0.00 ± 0.00% (0 ± 0 s)	0.00 ± 0.00% (0 ± 0 s)	25.61 ± 5.28% (886 ± 182s)	0.00 ± 0.00% (0 ± 0 s)

5.4.2 UPCON SIGNALING IMPACT

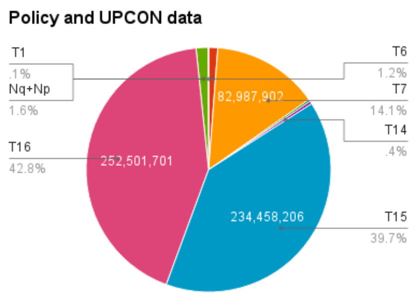
Another important metric that was analyzed was the signaling impact of the congestion reporting procedure in the operator's Control Path. The results on the amount of policy data, in bytes, exchanged for each trigger and the congestion reporting procedure (represented by *Nq+Np*), during the simulation for *Case_A* when each of the Policies 2, 3 and 4 are applied are presented in Figures 5.18 and 5.19. The event Triggers are defined in subsection 4.3.3. The results are summarized in Table 5.4.

For *Case_A*, in Figures 5.18a, 5.18b and 5.18c, the amount of data for congestion reporting is 1.2%, 1.6% and 1% of the total bytes for each of the policies respectively, and the policy data triggered by *Trigger_14* (T14 - Congestion status changes) represents under 0.4%. This is justified by the fact that the congestion signaling is, as seen in 5.4.1, almost limited to the lunch period and by the different content of both message types. The ANDSF policy corresponds to large MOs exchanged via the S14 interface in XML while the UPCON protocols exchange information on the number of users, their

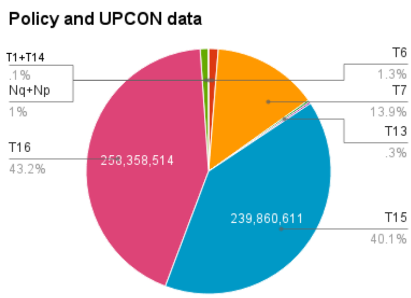
PDNs and the mobile Cell congestion status. Regarding the ANDSF Triggers, it is evident that the most significant amount of data is exchanged for *Trigger_15* (T15 - Wi-Fi AP lost) and *Trigger_16* (T16 - Wi-Fi AP found) followed by *Trigger_7* (T7 - change of mobility) with around 40%, 43% and 14% respectively in the three policies. Although with similar percentages, the total amount in bytes varies with the application of the different polices, comparatively with the values for *Policy_2*, T16 is 6.05% lower and T15 is 6.67% lower with *Policy_3*. Again, comparing with *Policy_2*, T16 is 3.87% lower and T15 is 4.52% lower with *Policy_4*. Comparing with Figure 4.7b with the application of *Policy_1*, the results are around 1% lower in *Trigger_15* (T15) and *Trigger_16* albeit higher in *Trigger_7* (T7). Regarding the amount in bytes, comparing with *Policy_1*, T16 and T15 are 3.90% 2.15% lower respectively with *Policy_2*, and with *Policy_3* T16 is 8% lower and T15 is 8.68% lower. With *Policy_4* T16 and T15 are 5.86% and 6.57% lower respectively.



(a) *Case_A Policy_2*



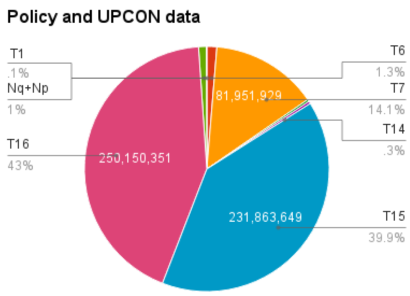
(b) *Case_A Policy_3*



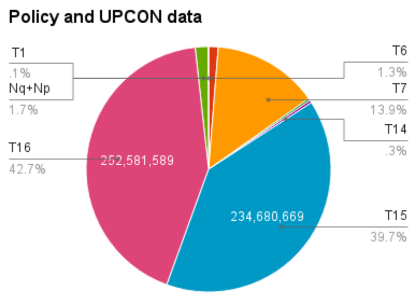
(c) *Case_A Policy_4*

Figure 5.18: UPCON Signaling Impact in *Case_A*

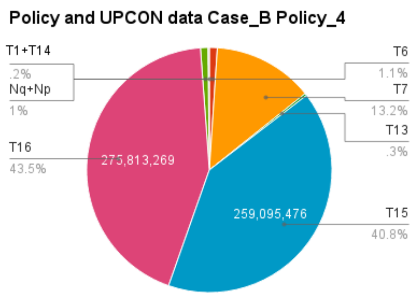
Figure 5.19 shows that for the amount of data regarding congestion reporting is 1% for *Policy_2*, 1.7% for *Policy_3* and 1% *Policy_4* of the total bytes. The ANDSF policy data triggered by *Trigger_14* is just 0.3% of the total bytes exchanged. Regarding the other triggers, it is noticeable that the most significant amount of data is triggered by *Trigger_15* and *Trigger_16*, followed by *Trigger_7* (T7) with around 40%, 43% and utmost 14% respectively in the three policies. Comparing with the results in Figure 5.18, the amount of data for the congestion reporting procedure is similar and around 1% versus the data for the ANDSF policy provision. Performing a comparison with Figure 4.7b where the results of the application of *Policy_1* are presented, the policy data is 1% lower in *Trigger_15* and *Trigger_16* albeit 1% higher in *Trigger_7* (T7). As in Figure 5.18, the total amount in bytes varies with the application of each policy. Comparatively with the values for *Policy_2*, T16 is 0.97% higher and T15 is 1.21% higher with *Policy_3*. Again, comparing with *Policy_2*, T16 is 10.26% higher and T15 is 11.74% higher with *Policy_4*. Performing a comparison with Figure 4.7b which presents the results for the application of *Policy_1*, the results are around 1% lower in *Trigger_15* (T15) and *Trigger_16* albeit higher in *Trigger_7* (T7). Regarding the amount in bytes, comparing with *Policy_1*, T16 and T15 are 8.86% 6.69% lower respectively with *Policy_2*, and with *Policy_3* T16 is 7.97% lower and T15 is 8.57% lower. With *Policy_4* T16 and T15 are 0.50% and 0.92% higher respectively.



(a) *Case_B Policy_2*



(b) *Case_B Policy_3*



(c) *Case_B Policy_4*

Figure 5.19: UPCON Signaling Impact in *Case_B*

Table 5.4 presents a summary of the results for the impact of the UPCON congestion report signaling (and each of the Triggers) in the overall amount of bytes exchanged in the simulation. For each Use Case and Policy, the relative percentages and the total amount of bytes are presented.

Table 5.4: UPCON signaling impact in the simulation

	Case_A				Case_B			
	Policy_1	*Policy_2*	*Policy_3*	*Policy_4*	*Policy_1*	*Policy_2*	*Policy_3*	*Policy_4*
UPCON	—	1.25%	1.57%	1.03%	—	1.01%	1.69%	0.98%
T1	0.06%	0.06%	0.07%	0.06%	0.07%	0.07%	0.07%	0.06%
T6	1.16%	1.21%	1.21%	1.28%	1.29%	1.31%	1.27%	1.05%
T7	13.34%	13.43%	14.05%	13.88%	14.08%	14.09%	13.94%	13.22%
T13	0.27%	0.29%	0.29%	0.27%	0.27%	0.29%	0.30%	0.29%
T14	—	0.27%	0.36%	0.24%	—	0.32%	0.35%	0.12%
T15	41.17%	40.34%	39.70%	40.07%	40.54%	39.88%	39.68%	40.82%
T16	44.01%	43.16%	42.75%	43.16%	43.76%	43.02%	42.71%	43.45%
Total bytes	623,628,122	622,695,848	590,636,605	598,557,272	587,723,350	581,458,611	591,428,448	634,711,296

From these results it is possible to infer that, in terms of signaling, the impact of the implementation of the congestion detection and reporting mechanism is not significant when compared with the whole ANDSF policy data exchanged during the simulation in either Policy 1, 2, 3 or 4. This reduced signaling for the RUCI reporting procedure must take in consideration the assumptions regarding the messages in subsection 5.3.1, i.e. the messages are exchanged without errors or optional AVPs, the RCAF sends information requests only to the MME (NqAP-IMSI-APN-INFORMATION-REQUEST and NqAP-IMSI-APN-INFORMATION-RESPONSE messages), RUCI reports are of the type Aggregated (ARR and ARA messages) and the variable information containing the number of agents that caused congestion was considered. Although some differences are verified in terms of the whole policy bytes exchanged for each Trigger, no significative discrepancies are verified in terms of signaling data with the application of the different policies. Therefore, comparing with the results for *Policy_1* (the policy defined in the previous existing work which implemented the ANDSF mechanism not taking into account congestion information), the implementation of the RUCI reporting mechanism has no relevant impact in the Network Control Path for either of the Policies which consider congestion information at the mobile Cell (Policies 2, 3 and 4).

5.5 CHAPTER CONSIDERATIONS

In this chapter the main practical work of this dissertation was presented. The network architecture of the proposal to enhance the ANDSF mechanism implemented in a previous existing work with a congestion reporting mechanism was detailed. The chapter concluded with the analysis of the simulation results on the two main Use Cases for the four Policies (ANDSF decision trees) that were implement in the simulation environment. The results reveal a significant decrease in the poor Wi-Fi service times up to zero bad service for both SSIDs in mall (when *Policy_4* is implemented with a Wi-Fi SSID supporting all the agents in the simulation), with minimal signaling impact on the operator network.

CHAPTER 6

Conclusion

This dissertation aimed at studying, leveraging and applying improved congestion management mechanisms in 3GPP networks. With the continuous growth of the number of customers with mobile data subscriptions, and the devices used to connect to the network and allowing a high-speed Internet access, the burden in the network infrastructure is increasing leading to poor user experience. This can be mitigated with costly infrastructure extension or with network optimization. Therefore, these technologies have been growing in importance for MNOs. Furthermore, research by Cisco [3] and Ericsson [1] predicts that this trend will continue with the upcoming introduction of 5G networks so research in this field is essential for current and future mobile networks.

Chapter 2 presented the state of the art on congestion management in mobile networks, starting with its drivers and the entities responsible for the industry standards. The focus was placed on 3GPP's technologies namely LTE, its Access and Core Network architecture, along with the relevant nodes being briefly described. Furthermore, an overview of several traffic mobility and offloading 3GPP mechanisms, such as the ANDSF, was performed. It was noted that the User Plane of the LTE architecture is the bottleneck as it serves all the traffic sent from and to the UE.

The recently standardized UPCON feature was overviewed in detail in Chapter 3. It was also noted that the standardized solution only involves Core Network elements, in either the congestion reporting or management, and that such implementation leaved room to enhancement and interworking with other solutions to achieve a wider congestion management solution.

In Chapter 4, an existing simulation environment developed to evaluate the deployment of ANDSF by a MNO was thoroughly analyzed, presenting its results. Enhancements to that work were suggested with the goal of implementing a congestion monitoring function, either for the 3GPP Cell or the Wi-Fi APs.

An architecture was proposed in Chapter 5, where the UPCON network function inter-works with the ANDSF. This solution improves the ANDSF mechanism as it enables the use of traffic offloading to Wi-Fi APs responding to the current level of network congestion. The proposed enhancements were implemented in the simulation environment and used to evaluate this solution.

The results reveal a significant improvement in general user experience, and according to the selected policy, permits zero bad service to Premium users, all with minimal impact on the operator infrastructure.

6.1 FUTURE WORK

While the implemented simulation environment achieved the goal of extending previous work to enhance an ANDSF rules engine with 3G/LTE cell congestion information and is capable of evaluating congestion both at the cell and at the Wi-Fi AP, improvements and optimizations can be introduced both at the ANDSF server and client interaction and at the UPCON cell network congestion reporting. Some key improvements are presented below:

UPCON:

- Cell reporting: extend reporting to all cells in the network;
- Reporting details: report at congestion levels and types of user traffic.

ANDSF:

- Rules engine: rules shall respond to the type of user traffic;
- Triggers: send rules only when the cell congestion changes.

References

[1] Ericsson, *Ericsson Mobility Report 2015*, May 2015. [Online]. Available: http://www.ericsson.com/mobility-report.

[2] L. Wade, *Where Cellular Networks Don't Exist, People Are Building Their Own*, Jan. 2015. [Online]. Available: http://www.wired.com/2015/01/diy-cellular-phone-networks-mexico/ (visited on 04/26/2016).

[3] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020", White Paper, Feb. 2016. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf.

[4] GSMA, *The Mobile Economy 2015*, 2015. [Online]. Available: http://www.gsmamobileeconomy.com/GSMA_Global_Mobile_Economy_Report_2015.pdf (visited on 06/07/2016).

[5] NTT DOCOMO, *Corporate Vision: A Story*, Jan. 2013. [Online]. Available: https://www.youtube.com/watch?v=LR3WPRinpn0 (visited on 04/22/2016).

[6] J. Kellokoski, J. Koskinen, and T. Hämäläinen, "Always Best Connected Heterogeneous Network Concept", en, *Wireless Pers Commun*, vol. 75, no. 1, pp. 63–80, Mar. 2014, ISSN: 0929-6212, 1572-834X. DOI: 10.1007/s11277-013-1348-9. [Online]. Available: http://link.springer.com/article/10.1007/s11277-013-1348-9.

[7] I. Demydov, M. Seliuchenko, M. Beshley, and M. Brych, "Mobility management and vertical handover decision in an always best connected heterogeneous network", in *The Experience of Designing and Application of CAD Systems in Microelectronics*, Feb. 2015, pp. 103–105. DOI: 10.1109/CADSM.2015.7230808.

[8] *3GPP specifications: Active work programme - UPCON*. [Online]. Available: http://www.3gpp.org/DynaReport/FeatureOrStudyItemFile-570029.htm (visited on 04/26/2016).

[9] Gartner, *Gartner Says Global Smartphone Sales to Only Grow 7 Per Cent in 2016*, Mar. 2016. [Online]. Available: http://www.gartner.com/newsroom/id/3270418 (visited on 05/19/2016).

[10] NTIA, *Evolving Technologies Change the Nature of Internet Use*, Apr. 2016. [Online]. Available: https://www.ntia.doc.gov/blog/2016/evolving-technologies-change-nature-internet-use (visited on 04/21/2016).

[11] Ericsson, *Ericsson Mobility Report MWC Edition February 2016*, Feb. 2016. [Online]. Available: http://www.ericsson.com/res/docs/2016/mobility-report/ericsson-mobility-report-feb-2016-interim.pdf.

[12] *Ericsson Mobility Report 2014*, Nov. 2014. [Online]. Available: http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf.

[13] B. Kramer, *How live-streaming is going to crush it in 2016*, und, Feb. 2016. [Online]. Available: http://www.socialmediatoday.com/social-networks/how-live-streaming-going-crush-it-2016 (visited on 05/06/2016).

[14] S. Research, *THE ESSENTIALS OF INTELLECTUAL PROPERTY, FROM 3G THROUGH LTE RELEASE 12*, May 2015. [Online]. Available: https://www.ericsson.com/res/docs/2015/ericsson-3gpp-submission-study-whitepaper-may-2015.pdf.

[15] Ofcom, *Ofcom announces winners of the 4G mobile auction*, Feb. 2013. [Online]. Available: http://media.ofcom.org.uk/news/2013/winners-of-the-4g-mobile-auction/ (visited on 05/27/2016).

[16] G. Daniels, *Turkey exceeds expectations with 4G spectrum auction, having backed away from 5G*, Aug. 2015. [Online]. Available: http://www.telecomtv.com/articles/spectrum/turkey-exceeds-expectations-with-4g-spectrum-auction-having-backed-away-from-5g-12780/ (visited on 05/27/2016).

[17] Telcordia, *Preventing a Growing Mobile Network from Becoming a CAPEX/OPEX Drain*, 2009. [Online]. Available: https://www.ericsson.com/res/ourportfolio/pdf/telcordia_white_papers/MC-COR-WP-020v2.pdf.

[18] S. C. Forum, "Small cell deployments Market status report | May 2016", Tech. Rep., May 2016. [Online]. Available: http://scf.io/en/documents/050_-_Market_status_report_-_Mobile_Experts.php.

[19] Ruckus, *Carrier-Class Wi-Fi*, Jan. 2015. [Online]. Available: http://a030f85c1e25003d7609-b98377aee968aad08453374eb1df3398.r40.cf2.rackcdn.com/wp/wp-carrier-class.pdf.

[20] ITU-T, *Recommendation I.130 - Method for the characterization of telecommunication services supported by an ISDN and network capabilities of an ISDN*, Jun. 1989. [Online]. Available: https://www.itu.int/rec/T-REC-I.130-198811-I/en.

[21] N. Networks, *LTE-Advanced Pro (Video)*, May 2016. [Online]. Available: https://networks.nokia.com/videos/video-lte-advanced-pro (visited on 08/14/2016).

[22] S. Sesia, I. Toufik, and M. Baker, *LTE – The UMTS Long Term Evolution From Theory to Practice*, 2nd. WILEY, 2011, ISBN: 978-0-470-66025-6.

[23] Keysight Technologies, *LTE-Advanced: Technology and Test Challenges*, Jul. 2014. [Online]. Available: http://literature.cdn.keysight.com/litweb/pdf/5990-6706EN.pdf.

[24] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for Mobile Broadband*, 2nd. United Kingdom, 2014, ISBN: 978-0-12-419985-9.

[25] *3GPP TS 23.002 V13.7.0 Network architecture (Release 13)*, Jun. 2015. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.002/23002-d70.zip.

[26] T. Ali-Yahiya, *Understanding LTE and Its Performance*, en. New York, NY: Springer New York, 2011, ISBN: 978-1-4419-6456-4. [Online]. Available: http://link.springer.com/10.1007/978-1-4419-6457-1 (visited on 10/24/2016).

[27] All about Wired and Wireless Technology, *3GPP Long Term Evolution (LTE): Quality of Service (QoS) in LTE*, Jan. 2013. [Online]. Available: http://4g-lte-world.blogspot.com/2013/01/quality-of-service-qos-in-lte.html (visited on 05/06/2016).

[28] *3GPP TS 23.203 V13.7.0 Policy and charging control architecture (Release 13)*, Mar. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.203/23203-d70.zip.

[29] Joe Deu-Ngoc, *System Architecture Evolution*, en, Dec. 2014. [Online]. Available: https://en.wikipedia.org/w/index.php?title=System_Architecture_Evolution%5C&oldid=740859434 (visited on 09/11/2016).

[30] *3GPP TS 24.302 V13.6.0 Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks; Stage 3 (Release 13)*, Jun. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/24_series/24.302/24302-d60.zip.

[31] *3GPP TS 24.312 V13.3.0 Access Network Discovery and Selection Function (ANDSF) Management Object (MO) (Release 13)*, Jun. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/24_series/24.312/24312-d30.zip.

[32] Ruckus, "How Interworking Works: A Detailed Look at 802.11u and Hotspot 2.0 Mechanisms", Tech. Rep., Jul. 2013. [Online]. Available: http://a030f85c1e25003d7609-b98377aee968aad08453374eb1df3398.r40.cf2.rackcdn.com/wp/wp-how-interworking-works.pdf (visited on 04/22/2016).

[33] *Wi-Fi CERTIFIED Passpoint | Wi-Fi Alliance*. [Online]. Available: http://www.wi-fi.org/discover-wi-fi/wi-fi-certified-passpoint (visited on 04/22/2016).

[34] W. Paper, *Hotspot 2.0 and ANDSF for Smart Mobile User Connectivity*, May 2014. [Online]. Available: http://www.alticelabs.com/content/WP-Hotspot2-0-and-ANDSF-for-Smart-Mobile-User-Connectivity.pdf (visited on 06/09/2016).

[35] *3GPP TS 23.401 V13.6.1 General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access*, Mar. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.401/23401-d61.zip.

[36] *3GPP TR 23.829 V10.0.1 Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO) (Release 10)*, Oct. 2011. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.829/23829-a01.zip (visited on 09/21/2016).

[37] R. Gupta and N. Rastoji, *LTE Advanced – LIPA and SIPTO White Paper*, 2012. [Online]. Available: https://www.aricent.com/whitepapers/lte-advanced-lipa-and-sipto (visited on 12/13/2016).

[38] *3GPP TR 23.859 V12.0.1 Local IP access (LIPA) mobility and Selected IP Traffic Offload (SIPTO) at the local network (Release 12)*, Apr. 2013. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.859/23859-c01.zip.

[39] *3GPP TR 23.861 V13.0.0 Network based IP flow mobility (Release 13)*, Jun. 2015. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.861/23861-d00.zip.

[40] *3GPP TS 23.261 V13.0.0 IP flow mobility and seamless Wireless Local Area Network (WLAN) offload; Stage 2*, Dec. 2015. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.261/23261-d00.zip.

[41] H. S. Soliman, *Mobile IPv6 Support for Dual Stack Hosts and Routers*. [Online]. Available: https://tools.ietf.org/html/rfc5555 (visited on 09/21/2016).

[42] *3GPP TS 23.161 V13.2.0 Network-Based IP Flow Mobility (NBIFOM); Stage 2 (Release 13)*, Mar. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.161/23161-d20.zip.

[43] *3GPP TR 22.805 V12.1.0 Feasibility study on user plane congestion management*, Dec. 2012. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/22_series/22.805/22805-c10.zip.

[44] *3GPP TS 22.101 V13.8.0 Service aspects; Service principles (Release 13)*, Mar. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/22_series/22.101/22101-d80.zip.

[45] *3GPP TR 23.705 V13.0.0 Study on system enhancements for user plane congestion management (Release 13)*, Dec. 2014. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.705/23705-d00.zip.

[46] *3GPP SA2 SP-64 Meeting Report.* [Online]. Available: http://www.3gpp.org/ftp/tsg_sa/tsg_sa/TSGS_64/Report/SA_64_Approved_Rep_v101.zip (visited on 09/21/2016).

[47] *3GPP TS 29.217 V13.3.0 Policy and Charging Control (PCC); Congestion reporting over Np reference point (Release 13)*, Mar. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/29_series/29.217/29217-d30.zip.

[48] *3GPP TS 29.405 V13.1.0 Nq and Nq' Application Protocol (Nq-AP); Stage 3 (Release 13)*, Dec. 2015. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/29_series/29.405/29405-d10.zip.

[49] *3GPP TS 29.213 V13.4.0 Policy and charging control signalling flows and Quality of Service (QoS) parameter mapping,* Mar. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/29_series/29.213/29213-d50.zip.

[50] *3GPP TS 29.303 V13.3.0 Domain Name System Procedures; Stage 3*, Mar. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/29_series/29.303/29303-d30.zip.

[51] D. H. Crocker, *Augmented BNF for Syntax Specifications: ABNF*, Nov. 1997. [Online]. Available: https://tools.ietf.org/html/rfc2234.

[52] J. Arkko, E. Guttman, P. R. Calhoun, and J. Loughney, *IETF RFC 3588 "diameter base protocol"(obsoleted by RFC 6733)*, Sep. 2003. [Online]. Available: https://tools.ietf.org/html/rfc3588.

[53] *3GPP TS 29.274 V13.6 3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C); Stage 3*, Jun. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/29_series/29.274/29274-d60.zip.

[54] *ITU-T Recommendation E.212 - The international identification plan for public networks and subscriptions*, Sep. 2016. [Online]. Available: https://www.itu.int/rec/T-REC-E.212/en (visited on 12/11/2016).

[55] *3GPP TS 23.060 V13.6.0 General Packet Radio Service (GPRS); Service description; Stage 2*, Mar. 2016. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.060/23060-d60.zip.

[56] R. Stewart, *Stream Control Transmission Protocol (SCTP)*, Sep. 2007. [Online]. Available: https://tools.ietf.org/html/rfc4960.

[57] P. Mockapetris, *DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION*, Nov. 1987. [Online]. Available: https://www.ietf.org/rfc/rfc1035.txt.

[58] B. Bertenyi, *Eurescom - 3GPP system standards heading into the 5G era*, Jan. 2014. [Online]. Available: https://www.eurescom.eu/news-and-events/eurescommessage/eurescom-message-1-2014/3gpp-system-standards-heading-into-the-5g-era.html (visited on 05/23/2016).

[59] H. Capelo and J. Cardoso, *Siafu-IM3W - Simulator for Congestion Management with Enhanced ANDSF*, Dec. 2016. [Online]. Available: https://gitlab.atnog.org/heldercapelo/siafu-im3w.

[60] Openet, *Openet Deploys World's First 3GPP RAN Congestion Awareness Function | Press Release*, Apr. 2016. [Online]. Available: http://bit.ly/2h0W537 (visited on 05/28/2016).

[61] J. Arkko, G. Zorn, V. Fajardo, and J. Loughney, *IETF RFC 6733 "diameter base protocol "(updated by RFC 7075)*, Oct. 2012. [Online]. Available: https://tools.ietf.org/html/rfc6733.